University of Zurich[UZH]

Dr. Alberto Huertas, Jan von der Assen, Katharina O. E. Müller, Chao Feng, Daria Schumm, Weijie Niu, Thomas Grubl, Nasim Nezhadsistani, Ahmad Abtahi, Reza Abtahi, Anderson Rocha (Edts).

Internet Economics XVIII

TECHNICAL REPORT – No. IFI-2025.01

January 2025

University of Zurich
Department of Informatics (IFI)
Binzmühlestrasse 14, CH-8050 Zürich, Switzerland

ifi

B. Stiller et al. (Eds.): Internet Economics XVIII: IntEco Seminar
Technical Report No. IFI-2025.01, January 2025
Communication Systems Group (CSG)
Department of Informatics (IFI)
University of Zurich
Binzmühlestrasse 14, CH-8050 Zürich, Switzerland
URL: http://www.csg.uzh.ch/

Introduction

The Department of Informatics (IFI) of the University of Zurich, Switzerland works on research and teaching in the area of computer networks and communication systems. Communication systems include a wide range of topics and drive many research and development activities. Therefore, during the autumn term HS 2024 a new instance of the Internet Economics seminar has been prepared and students as well as supervisors worked on this topic.
Even today, Internet Economics are run rarely as a teaching unit. This observation seems to be a little in contrast to the fact that research on Internet Economics has been established as an important area in the center of technology and economics on networked environments. After some careful investigations it can be found that during the last ten years, the underlying communication technology applied for the Internet and the way electronic business transactions are performed on top of the network have changed. Although, a variety of support functionality has been developed for the Internet case, the core functionality of delivering data, bits, and bytes remained unchanged. Nevertheless, changes and updates occur with respect to the use, the application area, and the technology itself. Therefore, another review of a selected number of topics has been undertaken.

Seminar Operation

Based on well-developed experiences of former seminars, held in different academic environments, all interested students worked on an initially offered set of papers and book chapters. Those relate to the topic titles as presented in the Table of Content below. They prepared a written essay as a clearly focused presentation, an evaluation, and a summary of those topics. Each of these essays is included in this technical report as a separate section and allows for an overview of important areas of concern, sometimes business models in operation, and problems encountered.
In addition, every group of students prepared a slide presentation of approximately 45 minutes to present its findings and summaries to the audience of students attending the seminar and other interested students, research assistants, and professors. Following a general question and answer phase, a student-lead discussion debated open issues and critical statements with the audience.
Local IFI support for preparing talks, reports, and their preparation by students had been granted by Dr. Alberto Huertas, Jan von der Assen, Katharina O. E. Müller, Chao Feng, Daria Schumm, Weijie Niu, Thomas Grubl, Nasim Nezhadsistani, Ahmad Abtahi, Reza Abtahi, Anderson Rocha, and Burkhard Stiller. In particular, many thanks are addressed to Chao Feng for organizing the seminar and for their strong commitment on getting this technical report ready and quickly published. A larger number of pre-

presentation discussions have provided valuable insights in the emerging and moving field of communication systems, both for all groups of students and supervisors. Many thanks to all people contributing to the success of this event, which has happened in a lively group of highly motivated and technically qualified students and people.

Zurich, January 2025

Contents

Chapter 1

Front-running and MEV Attacks on the Ethereum Network

Szczepan Gurgul

Abstract: Maximal Extractable Value (MEV) refers to profit opportunities that arise due to inefficiencies in decentralized networks like Ethereum. Arbitrageurs including miners, validators, and bots exploit the ability to reorder transactions within blocks to gain financial advantage. Techniques such as front-running, back-running, and sandwiching are common, with these strategies leveraging technical vulnerabilities in the network to extract value. In this report, we are making a comprehensive overview of the MEV attacks from an economic point of view. After defining them in the technical terms, we will review its' various attributes such as its' profitability, risks, legal and ethical concerns among others.

Contents

1.1 The emersion of blockchain technology

In this part, we will review the fundamental technical aspects of blockchain technology, and its historical beginnings followed by the blockchain trilemma phenomenon that explicates possible difficulties in designing blockchain systems that might be a gateway to many potential system threats. Additionally, followed by an introduction to some common blockchain threats, a placing MEV attacks within a broader schema of blockchain-based attacks is shown.

1.1.1 Peer-to-Peer ledger overview

First of all, in simple terms, blockchain can be defined as an immutable digital record of transactions, which is based on the underlying concept of Distributed Ledger Technology (DLT). As the name "blockchain" implies - the records are stored in the "blocks" that have the transaction data (in the form of Merkle Tree), timestamp, and a cryptographic hash of the previous block. In each of the headers of a newly mined or minted block, there is the previous block hash field - containing the previously mentioned hash pointing to the preceding block. By linking the blocks in this way a form of backward-linked list structure is obtained, creating a continuous chain of blocks so-called blockchain, and ensuring integrity, ordering, and immutability of data. [12]
Historically speaking, blockchain technologies started to be tackled when a couple of scientists in the mid-1990s attempted to solve the problem of keeping digitalized data in a safe, secure, and immutable way. The first notable attempt to achieve so was described in 1991 by Stuart Haber and W. Scott Stornetta during the implementation of a system where document timestamps could not be tampered with, a year later authors incorporated Merkle Trees into the design, whereby representing data nodes as leaves - allowed proving specific transactions within a block without a need to download the entire block, this feature guaranteed efficient proof of transaction inclusion, improved lookup, and verification speed [10]. Many years later after another research and peer-to-peer (P2P) system development attempts in the year 2008 the anonymous person or group called Satoshi Nakamoto introduced the white paper - Bitcoin: A Peer-to-Peer Electronic Cash System, giving more public recognition and accelerating the adoption of decentralized digital currencies, and a broader concept of blockchain. [1]
One of the core structures of blockchains are consensus mechanisms that are used to bring all nodes of a P2P system to a common agreement based on some available data. Apart from validating transactions, consensus mechanisms secure the blockchains from many different types of blockchain network-based attacks.

- Proof of Work (PoW)

  For example, in Bitcoin, in the Proof of Work (PoW) consensus mechanism, miners solve complex cryptographic puzzles to validate transactions and append new blocks to the blockchain. The miner changes the nonce to guess the correct hash and mines the block receiving a block reward in Bitcoin (BTC) cryptocurrency [10]. The computational effort needed to solve a cryptographic puzzle makes it more difficult (and expensive - due to energy usage to run mining computations) for any single entity to try to alter the blockchain, increasing security and averting double spending problems.

- Proof of Stake (PoS)

  In the Proof of Stake (PoS) consensus mechanism, used in Ethereum 2.0 validators (block producers) validate new blocks to the blockchain based on the stake amount they lock up as collateral. For any malicious behaviors, validators are penalized

(slashed). Currently to become a validator 32 Ethereum cryptocurrencies (ETH) are needed, at the time of writing 1 ETH is worth 3200 USD, therefore to run one validator an entity needs to put almost 100,000 USD as collateral. The high cost of the addition of a new validating node makes the network secure economically rather than computationally, executing, for example, Sybil attacks (creating new malicious nodes) becomes inefficient. Also, since there is no solving of the cryptographic puzzle compared to PoW, energy usage is lower and the PoS consensus mechanism is considered to be more "eco-friendly".

On 30th July 2015, the Ethereum PoW network had its first genesis block marking the beginning of currently one of the most popular turing complete blockchains, allowing users to deploy programs called smart contracts onto the blockchain, and letting developers create decentralized applications for various cases. In this paper, we will analyze both Ethereum PoW and Ethereum 2.0 protocol with a Proof of Stake consensus mechanism (Ethereum Network following the Paris Upgrade on the 15th of September 2022), and how its architecture and design decisions undertaken empowered transaction front-running and different Maximum Extractable Value (MEV) attacks.

1.1.2 Navigating the blockchain trilemma: a gateway to potential threats

Before exploring Maximum Extractable Value (MEV) attacks it is important to understand the architecture design limitations of decentralized networks. When designing new Layer 1 or Layer 2 blockchains, certain design decisions taken might be a gateway to potential threats and defects in a system.

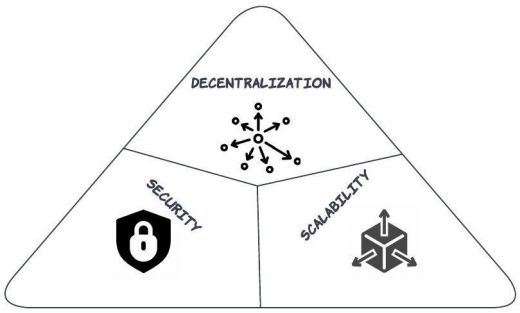


Figure 1.1: The Blockchain Trilemma

As stated by the Blockchain Trilemma Theory lodged by the Ethereum founder Vitalik Buterin, every blockchain faces a trade-off among key three design attributes: decentralization, security, and scalability. According to the trilemma, it is challenging or even impossible to optimize all three parameters simultaneously without compromising one of them [13]. Difficulties arise when one of the parameters is solemnly de-prioritized leading to many different types of blockchain threats that are going to be described below. Since, different attacks base on different balances, while de-prioritizing:

- Security - guides to blockchain security exploits, for example, smart-contract, finality, or centralized validators exploits. In a delegated proof of stake consensus mechanisms (dPoS), the goal is to quickly scale throughput (scalability) but with the sacrifice of decentralization, and having a small delegated number of validators (in an EOS chain for example 21 block producers). One of the biggest concerns still to this day, in dPoS, is that a sufficient number of "trusted validators" could be bribed by malicious actors to perform for example rollbacks of transactions.

- Decentralization - could lead to centralization-based attacks i.e. Sybil attacks, where the attacker creates numerous not honest identities to take control over the network and disrupt consensus. A historical example of this kind of incident could be the attack on Ethereum Classic (ETC) on the 5th of January 2019, where malicious attackers took control of a majority of computational power to validate the transactions and allowed themselves to double-spend 219,500 ETC . The attack happened due to the fact that Ethereum Classic (ETC) has a relatively small amount of validators (de-prioritizing decentralization) compared to Ethereum (ETH), for attackers it was less difficult to perform creation of fake identities to take over the network. Subsequently, after these and other incidents within the system, a hard fork of the Ethereum Classic happened, with the actuation of the Thanos Upgrade that allowed miners with 3GB and 4GB GPU to resume to mining ETC (due to change of epoch duration from 30,000 to 60,000 blocks), eventually allowing more miners to participate in the network, and therefore increasing network security. [21]

- Scalability - could lead to Front-running and MEV exploitation. For example, on the Ethereum blockchain - scalability has always been described as an ongoing issue. The average block creation time is around 12 seconds and the maximum throughput is recorded as 62 transactions per second (TPS) [9], combined with an open mempool for unconfirmed transactions allows users to observe and for example, do any kind of MEV exploitation which will be explained in later sections.

1.1.3 Types of blockchain threats

Generally, looking at blockchains from a cybersecurity perspective there are many threats to P2P networks like distributed denial of service attacks, routing attacks, or previously mentioned Sybil attacks. However, there are more risks than just trying to compromise the blockchain network directly. Blockchain users can also victimized more personally, through crafted user wallet attacks like phishing, where attackers through for example fake websites, emails, or messages are trying to trick users into connecting their personal wallets to malicious websites and subsequently approve the spending of some tokens - allowing attackers to drain up users wallets and effectively steal all of their funds.
Furthermore, another kind of blockchain attack are smart contracts attacks. Vulnerabilities in smart contract code can be exploited by attackers to bypass some functions and perform malicious activities. This kind of attack can be mitigated (not entirely excluded) by having smart contract audits where a team of trusted experts assesses the security and reliability of code, a superior example would be an audited open-source code of cryptocurrency HEX, a blockchain version of the certificate of deposit which allows users to stake their native HEX coins and earn yield upon it. All core logic is enclosed within one well-audited immutable smartcontract, and at the time of writing it has maintained 100 percent up-time with no known security exploits.
Another well-known category of attacks in blockchains are transaction verification mechanisms attacks, this kind of attacks specifically target the protocols that are responsible for the confirmation and verification of transactions. Although performing such a hack without having 51 percent computational power (hash rate) over block production is hard, it is not impossible. An example could be a double-spending attack that tries to alter the blockchain operations and allow the attacker to use the same input (cryptocurrency) more than once [11]. The nature of the public ledger of transactions in public permissionless blockchains like Bitcoin and Ethereum ensures transparency of transactions, thus if double-spending of the same input has happened it would be visible to everyone in the network, raising serious security doubts and loss of trust in the protocol. The attackers could use techniques like race-attacking based on an attempt to send two conflicting

transactions to distinct nodes hoping to exploit the delay in propagation of transactions in the network. Another worth mentioning transaction verification mechanism hack is called the "Finney attack" named after American software developer and early Bitcoin contributor Hal Finney [22]. The attack is based on having to privately pre-mine a block with a fraud transaction and then trying to spend the same funds somewhere else before the block is added to the blockchain, therefore creating two conflicting transactions. All in all, consensus mechanisms are helping to prevent double-spending of the same input, for example in PoW the computational effort needed to solve a cryptographic puzzle makes it more difficult for any single entity to try to alter the blockchain, averting double spending problems.
With all this insight, Maximum Extractable Value (MEV) attacks in a great schema of blockchain threats could be placed as a separate type of threat, somewhere between smart contract attacks and transaction verification mechanisms attacks. It is due to the fact that MEV does not compromise any kind of blockchain networks, user wallets, smart contract logic, or transaction ordering mechanisms but rather leverages already existing transactions and consensus protocols to extract profit from others. However, more on that in the following section.

1.2 Maximum extractable value (MEV) on a blockchain

In this part, the broader definition of MEV as well as its technical aspects and divergence from high-frequency trading (HFT) from traditional finance will be stated. The different MEV techniques like front running, back running and sandwich attacks will be introduced and thoroughly explained.

1.2.1 MEV definition and technical aspects

Overall, MEV refers to the maximum value block producers (miners/validators) can obtain by including, reordering, or excluding transactions when they produce new blocks, for example by prioritizing transactions with higher fees. Studies show that more than 95 percent of the miners or validators choose to order the transactions in descending order with respect to the gas price [3]
To understand the MEV attack, at first, it is important to understand where it comes from. MEV attack is conceptually similar to high-frequency trading (HFT) and draws some inspiration from it, both of the mechanisms exploit financial opportunities that arise from market inefficiencies, HFT operates in the scope of traditional finance (TradFi) whereas MEV occurs in the realm of decentralized finances (DeFi). HFT is an algorithmic trading method in which many orders are executed by hyper-speed complex algorithms used to execute orders depending on possible arbitrage opportunities on markets [14]. One positive aspect of HFT is that the large volume of arbitrage transactions improves market liquidity. MEV attacks similarly to HFT, but in the realm of decentralized finances improves liquidity between many different liquidity pools, giving better deals for traders often being harmed by price impact from trading on pools with lower liquidity.
Notably, the fundamental root cause behind enabling MEV attacks in blockchains is their mempool design and the presence of automated market makers (AMM) in DeFi.

- Automated Market Makers (AMM) Automated Market Makers (AMM) is a mechanisms that operate on liquidity pools on most major decentralized exchanges (DEX) like Uniswap, Curve, Sushiswap, and 1inch. They set up token prices based on the ratio of assets in the pool, in short, while swapping two tokens within a liquidity pool, the ratio of tokens in the liquidity pool changes, and AMM recalculates the ratio of assets to determine new price for each of the tokens. This kind of transaction

mechanism used by AMMs is far from ideal due to calculation inaccuracies since a large number of transactions would be needed to restore market-fair prices (for example, compared with other liquidity pools on different exchanges), concurrently opening up space for using MEV techniques to balance liquidity [15].

- Open mempool for unconfirmed transactions

  One of the key features of the public permission-less Layer-1 blockchains like Ethereum or Bitcoin is open mempool. The name mempool generally comes from "memory pool", which is a storage area where pending transactions are held temporarily before they are confirmed and then submitted to the blockchain. The whole flow of how a transaction is included in newly minted blocks on the Ethereum network is depicted in Figure 1.2. Open mempool means that it is free to analyze and look into by anyone interested, MEV attackers analyze mempool to spot the potential "victim" transactions and perform attacks, an attacker is usually encouraged to perform an attack if he can spot an arbitrage opportunity that will allow him to make some money. On the other hand side, most of the private permissioned blockchains for private projects will have closed mempools making it impossible for MEV attackers to tackle any of the transactions.

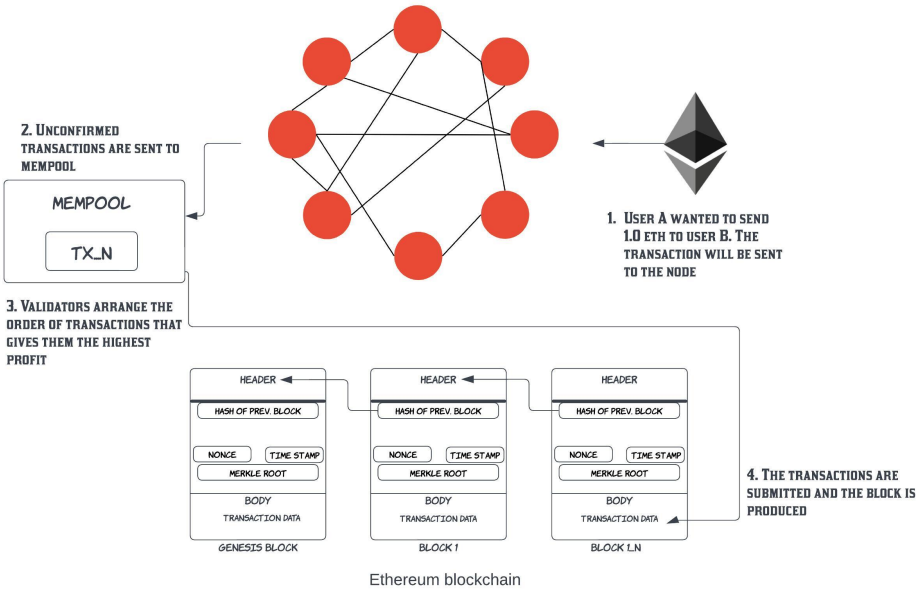


Figure 1.2: Ethereum blockchain transaction flow

1.2.2 MEV attacks

Overall, we can distinguish a few most common types of MEV attacks being performed on blockchain networks. These are: front-running attacks, back-running attacks, sandwich and bribery attacks

1.2.2.1 Bribery attacks

Bribery attacks generally refer to an attempt to bribe miners/validators to prioritize the order of transactions, usually done by monetary influence by setting high gas prices to incentivize block producers to include attacker transactions in order he wants to (validators will prioritize transactions with higher gas fees). Generally, bribes can be divided into two types: guided and effective bribing. [5]

- Guided bribing -

  The bribe is given as long as the bribed party behaves as instructed for specific instructions like reordering transactions within a mempool. Setting a higher transaction fee or gas fee by an attacker is form of a guided bribing since it incentivizes block producers to reshuffle the mempool but if the transaction was placed in bad order the miners/validators are not penalized.

- Effective bribing -

  The bribes are conditional on attack success, the payment of a bribe to the block producer is only given if and only if an attack has succeeded. Flash loans are a good example of effective bribing, In DeFi it is a unique feature that allows users to borrow capital without needing collateral, under the condition that the loan is repaid within the same block. This means that the MEV bot, performing an attack can ask for a loan, try to perform an attack and if the attack is successful, block the producer who included a fraudulent transaction will be paid, as well as borrow money will be repaid to the protocol. If the flash loan is not repaid within the same block, the transaction is reverted and the validator will not get any bribe.

From a blockchain network perspective, bribery attacks can be used to incentivize validators to participate in the double signing of blocks that could lead to potential double-spending of the same input and therefore create two conflicting transactions upon which the attacker will capitalize.

1.2.2.2 Front-running attack

The idea of front-running comes from traditional finance where if used based on insider knowledge and insider trading it is considered to be illegal. It is highly regulated by the Security Exchange Commission (SEC) in the USA and the Forestry Commission England (FCE) in the United Kingdom, as well as in other countries. Moreover, there are severe penalties for individuals and institutions that are caught in front-running, including fines and bans. Additionally, the front-running is widely used in HFT strategies. An example of illegal front running in a matter of traditional finance could be, a broker taking advantage of his "insider knowledge" of upcoming trades. Broker 'X' knows their client 'Y' will place a large trade (that will move the price up) and buy the same asset before the client's order is executed, profiting from the expected price movement that would be caused by the client's trade [4].
In the realm of blockchains, MEV bots are monitoring public mempool transactions and are looking for large trades that will be profitable for them. When the MEV bot spots an opportunity it will submit its own trade ahead of the victims' trade hoping for profit from the price movement. This exact procedure can be seen in figure 1.3. To achieve its effect of placing its trade before the victim's trade, the MEV attacker incentivizes validators to include his transaction by putting higher transaction fees compared to the victim fee. In that sense, on the Ethereum network, for validators it is more profitable (due to higher gas tip) to put attackers' transactions before the victim's (users) transaction, this concept is often referred to as bribing. [6]

1.2.2.3 Back-running attack

Back-running, opposite to front-running is based on placing a trade immediately after a large market-moving trade rather than before it. This exact procedure can be seen in figure 1.4. The bribing to position the attacker transaction behind the victim transaction can be done as well by manipulating the gas fees for the validators.

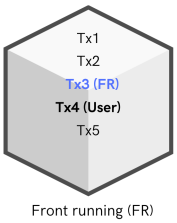


Figure 1.3: Front-running on Ethereum network

The positive aspect of the back-running attack is the ability to restore market imbalance, by quickly purchasing the asset after the price moves down, it works kind of as reactive liquidity that enters the market when the opportunity comes, rather than passive liquidity being deposited by liquidity providers to the liquidity pools.
An example could be cascade liquidations when one liquidation triggers the others and by chain reaction price falls down dramatically. This can happen when on DeFi for example too many loans are collateralized with leveraged positions therefore being sensitive to the price movement of an asset. When one position is liquidated this will cause a sell-off of collateral and push the price even lower making other open positions liquidate. MEV bots monitor open mempool, spot the opportunity and place back-running transactions to capitalize on the sell-off of this asset, this kind of provision of the reactive liquidity helps to absorb the sell-off impact and prevent further cascading.

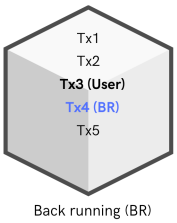


Figure 1.4: Back-running on Ethereum network

1.2.2.4 Sandwich attacks

Sandwich attacks are a combination of both front-running and back-running methods. The MEV bots monitor the mempool and then submit two transactions that ”sandwich” the targeted transaction. The first one is a front-running transaction that is placed before the victim’s transaction - used to buy an asset before its price is pushed up. The second one is a back-running transaction that after both of the previous transactions are included (two buy orders) will automatically sell an asset. This exact procedure can be seen in figure 1.5. The attack makes the victim buy an asset at a greater price (after the FR transaction), effectively extracting value from his purchase power and diminishing the amount of an asset that he would have gotten if he had not been attacked.
One fascinating fact is that for a long time the MEV bot performing mostly sandwich attacks with an address called ”jaredfromsubway.eth” has been one of the top gas spenders on the Ethereum network, the total adjusted value over time paid in bribes and transaction

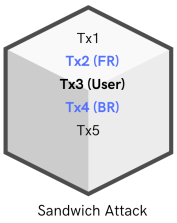


Figure 1.5: Sandwich attack on Ethereum network

fees over a course of time from March 2023 to November 2024 equals to 216,339,259 USD as a time of writing [16]. This enormous amount of money spent on transaction fees helped the MEV bot to perform thousands of MEV attacks a day with great profitability. An example of one of these attacks is from Ethereum block number: 17178637 [17]

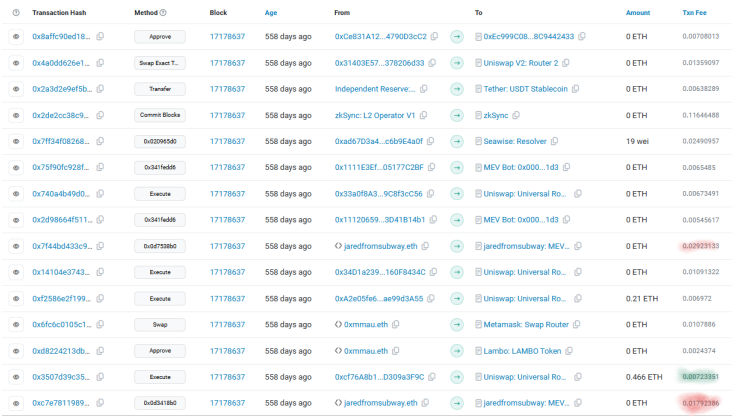


Figure 1.6: Example of MEV sandwich attack at block number '17178637' on Ethereum network

As visible in figure 1.6, the first transaction marked with a red color marker has been a front-running transaction performed by the bot jaredfromsubway.eth, the bot has set a higher transaction fee (column Txn Fee) 0.02923133 than the victim's transaction fee 0.00723351 to effectively incentivize validator to put his transaction first. Furthermore, after the user transaction had been placed in a block, he closed the sandwich attack with back running transaction in which he swapped the bought previously asset back to Ethereum (ETH), making around 0.02 ETH (70USD at the time of writing) as instant profit [17].

1.3 MEV attacks as cryptocurrency arbitrage opportunity

In this section, a list of suggested attributes is presented to assess the MEV attacks as a cryptocurrency arbitrage opportunity. Comprehensive analysis in the context of root cause, durability, capital intensity, MEV risks, and methods to avoid it are going to be presented. Moreover, ethical and unethical attributes are mentioned together with historical MEV returns.

1.3.1 Root cause

One of the causes of the MEV attacks on the Ethereum network lies in blockchain design elements, such as the open transparency of the mempool, and deterministic mechanisms in DeFi like automated market makers (AMM). Most often, to perform an MEV attack, the bots need to first detect the arbitrage opportunity within the mempool that they can capitalize on. There might be many different root causes for the arbitrage opportunities to arise. One of the examples could be:

- Different depth of liquidity pools

  If there are liquidity pools of different depths on separate exchanges, trading assets within a liquidity pool with a high price impact can create price disparities and arbitrage opportunities. If for example, some significant trade comes on a liquidity pool with lower liquidity, the price impact from this action will be greater than comparing price impact for having the same transaction size on a liquidity pool of bigger depth. This price impact made by the transaction (either buy or sell) will move the price significantly, the MEV bots would compare the price with the other exchange and spot this situation as an arbitrage opportunity, buying up the asset from the liquidity pool where the asset is cheaper and selling it in another place, capitalizing on the price difference.

- Oracle lag updates

  Lending protocols like Aave or dYdX DAOs use external Oracles (for example, Chainlink) to supply real-time off-chain data into blockchains. These data feeds provide up-to-date pricing of assets used by protocol (i.e. used liquidations or lendings) to make sure that these assets have the newest price with respect to the current market conditions. If Oracle lag happens, and asset price on lending protocol still uses outdated price, arbitrageur bot can detect price discrepancy between an outdated price and new price being available on the real-time market and exploit this lag, using for example outdated data for his advantage. This kind of behavior can lead to unfair situations within a blockchain where some loans might be liquidated without rational reason or some assets can be bought at inaccurate prices.

1.3.2 Durability and delta-neutrality of MEV

In the context of this paper, durability is an attribute that measures how long an arbitrage opportunity is expected to last before being neutralized by market forces, i.e. supply and demand dynamics [2]. MEV attacks are usually performed by bots. Regarding MEV attacks as arbitrage opportunities, durability is low since MEV bots exploit very short-lived inefficiencies created in mempools or by inaccuracies created by AMM in liquidity pools. Additionally, since many very competitive MEV bots are competing in an open mempool and each new block on the Ethereum network is produced every 12 seconds durability measured as arbitrage potential can be measured as almost instant.
The directionality of MEV attacks, or in other words delta-neutrality - indicates whether an arbitrage strategy relies on the directional movement of asset prices. Most of the MEV attacks use high-performance bots, whose strategies generally aim to extract value irrespective of falling or rising markets. MEV attacks as cryptocurrency arbitrage opportunities are transaction-focused, meaning that they rather look for a quick scalp of profitable transactions rather than being dependent on the price performance of an asset in the long term.

1.3.3 Capital intensity

The capital intensity generally refers to the need for significant capital to start attacks, capital-intensive strategies typically offer higher returns but also greater risks of losing invested capital. Generally speaking, MEV attacks are capital-intensive, the MEV bot has to be equipped with the chain native cryptocurrency to pay for the transactions executed, as well as for bribing the validators (i.e. while frontrunning). In lending protocols like Aave, if the collateral of users goes below some threshold he gets liquidated, The MEV bot can quickly step in to repay the debt and in return get users' collateral getting profit. This kind of process is very capital intensive since a substantial amount of money is needed to cover the loan amount needed to trigger liquidation.
In DeFi, there exist some features like flash loans, that allow users to borrow capital without needing collateral, if and only if the loan is repaid within the same block (reminder: on Ethereum, every new block on average is produced every 12 seconds). If a flash loan is not repaid within this period, the transaction is reverted. This application is a perfect fit for MEV bots performing MEV attacks since they do not need the upfront capital, and after a successful attack loan can be instantly repaid. When taking a flash loan, the additional fees for protocol providers (like Aave - 0.09 percent of the loan) have to be paid.

1.3.4 On-chain and off-chain analysis

In MEV attacks on the Ethereum network, all arbitrage operations are performed on-chain. This kind of on-chain arbitrage brings certain advantages and disadvantages for the MEV attackers as well as other participants in the market, mainly:

- On-chain arbitrage advantages

  One of the advantages of MEV attacks on public permissionless networks like Ethereum is the transparency of transactions that are crafted to perform the attacks, allowing external verification of data and pattern recognition for researchers. Furthermore, thanks to arbitrageurs liquidity pools are more stabilized and users can enjoy reduced slippage and therefore better efficiency of their capital. Another positive aspect of MEV arbitrage is reducing price indifference between DEXs giving more consistent pricing of assets across multiple markets. Another positive aspect might be that validators due to MEV attacks are getting additional gas tips, effectively boosting their ROI from running a validator, this can incentivize more block producers to join the network and therefore make it more decentralized.

- On-chain arbitrage disadvantages

  Due to many MEV bots, often daily executing hundreds of arbitrage transactions and MEV attacks network congestion rises. At the same time, gas prices spike, and ordinary users are harmed by increased gas fees on the Ethereum network. Most of the MEV attacks, from front running to sandwich attacks, aim for opportunities and for potential victims submitting transactions to the public mempool. This means that any unaware user of DeFi can be a victim of an MEV attack and his purchase value can be extracted from his trades, effectively harming him.

1.3.5 MEV risks

If one runs an MEV bot to perform MEV attacks, the bot can encounter some risks and traps like honeypot smartcontracts. The name honeypot comes from the area of cybersecurity, where the honeypot system acts like a decoy specifically designed to attract

attackers in order to trap them and gain some valuable piece of information. In the context of the Ethereum blockchain, honeypot smartcontracts are made to lure MEV bots into faulty design smartcontracts that have functions to effectively drain or lock Bots capital. For example, built-in smartcontract functions by design can allow only certain parties to interact with it, for example - enabling only the contract deployer access to sell tokens effectively blocks other users from selling and therefore locks their capitals if they have traded a given token in a liquidity pool (or in other cases for poison token smartcontracts, approved the spending of a token). The fake buys are orchestrated by the team behind the honeypot smartcontract to give the impression of a robust and fair trading market and attract MEV bots to the honeypot smartcontracts by displaying some arbitrage opportunities. In this case, if MEV bots interact with the honeypot, its funds are locked or drained.
On the other hand, looking from the DeFi users' perspective, there exists a way to mitigate being a victim of an MEV attack. One suitable solution would be using DEX aggregator CowSwap DAO, which has an anti-mev attack system utilizing batch auctions. The idea is very similar to how sequencers work on Layer-2 scaling solutions like zkSync or Arbitrum, where a mechanism bundles many transactions into batches (for example, 1 batch = 100 transactions) and executes them in batch. This approach hinders MEV bots from spotting an arbitrage opportunity (since singular transactions and possible arbitrage opportunities cannot be distinguished from other transactions included in a batch by Bots) and therefore protects users from being victims of an MEV attack.
Additionally, to protect themselves from MEV attacks, DeFi users can use Flashbots DAO and their "Protect RPC" to directly submit their transactions to the validators, without expose to a public mempool where they can be victim of an MEV attack. Flashbots DAO has as well functionality tailor-made for Ethereum validators, MEV-Geth, a modified version of the Ethereum client, allowing validators to receive bundles of transactions directly from traders at the same time boosting their profitability by capturing MEV rewards.

1.3.6 Ethical and unethical attributes

Overall, many people could argue if MEV attacks are considered to be ethical or unethical, also projecting onto the legal and illegal aspects of performing this kind of attack. In general, while most people see MEV attacks as unethical (malicious) actions, others see them more as ethical arbitrage opportunities due to their built-in blockchain technical aspects. Therefore let us analyze them in those two dimensions:

- Unethical and illegal aspects

  MEV attacks as a way to perform cryptocurrency arbitrage can be considered unethical or illegal since they harm and extract value from common DeFi users. The potential victim being frontrunned has its purchase power lowered by getting a worse deal within the liquidity pool, additionally, in this case, other transactions can be excluded from the new block, making it unfair and unethical toward other network participants. Due to MEV attacks, every DeFi user is also suffering from increased gas fees needed to perform any on-chain operations, which effectively could be considered unethical since it harms users who already decided to submit any transactions, making them speed up transactions by adding new gas (spending additional capital) or have their transactions delayed for some time.

- Ethical and legal aspects

  If looking at traditional markets for reference, frontrunning using insider information is illegal and penalized. However, frontrunning in the realm of blockchains, specifi-

cally Ethereum happens in an open mempool of transactions that is permissionless, meaning everyone can access it, therefore there is no "insider information" involved. Moreover, in pure P2P networks, there is no central authority that can use its power to get some additional information that can be used in frontrunning, making MEV attacks lean more towards ethical or legal actions rather than unethical or illegal ones. Another aspect is the idea of "code is law" stating that whatever code (i.e. smartcontract) permits then it is "legal" - assenting with American "First Amendment law" including freedom of speech and writing. Some people could also agree that MEV attacks are ethical since they fill the gap in the pricing of an asset created by the inaccuracy of AMM. The bot by doing cryptocurrency arbitrages evens the prices in two different locations and makes them consistent across markets.

Knowing all of these aspects and different approaches to understanding these attacks as arbitrage opportunities, it is very challenging to justify MEV attacks and assess their ethics and legality. Unless some party decides to regulate the cryptocurrency market more, MEV attacks can be also considered but not sentenced to be illegal.

1.3.7 Historical returns

The last attribute to analyze when considering MEV attacks as cryptocurrency arbitrage opportunities are monetary returns that incentivize more and more MEV Bots to participate and compete in creating sophisticated attacks.
Overall, validators running nodes on the Ethereum network often also run MEV bots aside from typical node operations, leveraging a node's close proximity to mempool and control over transaction processing and ordering. Therefore, aside from having around 3.5 percent APR and tips from priority fees, validators can significantly boost their revenue by performing various MEV attacks.
In previous sections, we have discussed the Ethereum address called: jaredfromsubway.eth being responsible for a substantial amount of transaction traffic on the Ethereum blockchain and at the same time generating thousands of profitable MEV attacks each day. To fully analyze the monetary returns of this address from performing MEV attacks, on-chain analysis has to be performed. Luckily, websites like Eighen-phi [18] provide insights into transaction structures, focusing on many arbitrage strategies and MEV attacks and enabling researchers and DeFi members to a real-time on-chain profitability analysis of different MEV attacks. After examining the website in more detail, an interesting MEV bot has been spotted of address '0x1f2F10D1C40777AE1Da742455c65828FF36Df387' [20]. At the time of writing following the MEV bot, for 7 days performed 16,000 MEV attacks were performed, spending a total of 2.326 mln USD in gas fees, generating 2.355 mln USD of revenue, and therefore around 30,000 USD of profit. On the other side of the trench, address like '0x3e28b1d60a47eD10Fa1025d35d772589d6698C0b' [19] have been a victim of a sandwich attack and their losses exceeded 90,000 USD. As we can see, MEV attacks can be very profitable for attackers, leading private developers and institutions to develop better and better bots to outcompete rivals in a public mempool and capture better cryptocurrency arbitrage opportunities.

1.4 Summary

This paper thoroughly investigated the phenomenon of maximal extractable value (MEV) on a decentralized network like Ethereum, emphasising the technical aspects of blockchains and the financial incentives behind attacks as cryptocurrency arbitrage opportunities.
The study commenced with a practical introduction to peer-to-peer (P2P) systems, followed by an introduction to the blockchain trilemma phenomenon. Together, these two

sections made the reader acquainted with underlying concepts of blockchains like distributed ledger technologies, and their architectural design trade-offs, which often serve as a gateway to subsequently explained potential threats to the blockchain and its users. Furthermore, a broader definition of MEV was established within the context of high-frequency trading (HFT) and traditional finance. Various MEV attack techniques, such as front-running, back-running, sandwich, and bribery attacks, were explained in detail, providing the reader with a comprehensive understanding of these concepts in the realm of blockchains.

Finally, the financial incentives behind MEV attacks as cryptocurrency arbitrage opportunities were analyzed, focusing on various attributes such as capital intensities or historical returns. Additionally, the paper thoroughly explored the root causes and market durabilities of arbitrage opportunities, along with the risks associated with running MEV bots and strategies to mitigate the value extraction for DeFi users. Building on this foundation, the ethical and unethical aspects of MEV attacks were examined, encouraging further consideration of the topic by readers.

In conclusion, MEV and Frontrunning attacks on the Ethereum network highlight complex trade-offs between financial arbitrage opportunities and the challenges of blockchain ecosystems. On the one hand, enhancing market efficiencies by improving liquidity distribution across pools, but on the other exploiting users placing trades within decentralized finances and raising ethical concerns. Balancing these aspects is crucial to ensure sustainable, long-term focused growth of blockchain ecosystems and calls for further research.

Bibliography

[1] Nakamoto, S. (2008) *Bitcoin: A Peer-to-Peer Electronic Cash System*, Retrieved November 11, 2024, from `https://bitcoin.org/bitcoin.pdf`.

[2] Abtahi, A., Abtahi, R. (2024): *A taxonomy of inefficiencies and arbitrage opportunities in cryptocurrency markets. Working Paper, UZH.*

[3] Varun, M., Palanisamy, B., Sural, S. (2022, May): *Mitigating frontrunning attacks in ethereum. In proceedings of the Fourth ACM International Symposium on Blockchain and Secure Critical Infrastructure*, (pp. 115-124).

[4] Torres, C. F., Camino, R. (2021): *Frontrunner jones and the raiders of the dark forest: An empirical study of frontrunning on the Ethereum blockchain*, In 30th USENIX Security Symposium (USENIX Security 21) (pp. 1343-1359).

[5] Karakostas, D., Kiayias, A., Zacharias, T. (2024, February): *Blockchain Bribing Attacks and the Efficacy of Counterincentives.*

[6] Coinmonks (2023): *Frontrunning: Understanding MEV Attacks*, Retrieved September 28, 2024, from `https://medium.com/coinmonks/frontrunning-understanding-mev-attacks-406df02d8bb5`.

[7] Eskandari, S., Moosavi, S., Clark, J. (2020): *Sok: Transparent dishonesty: frontrunning attacks on blockchain. In Financial Cryptography and Data Security: FC 2019 International Workshops, VOTING and WTSC,*, St. Kitts, St. Kitts and Nevis, February 18–22, 2019, Revised Selected Papers 23 (pp. 170-189). Springer International Publishing.

[8] Daian, P., Goldfeder, S., Kell, T., Li, Y., Zhao, X., Bentov, I., ... Juels, A. (2020, May) *Flash boys 2.0: Frontrunning in decentralized exchanges, miner extractable value, and consensus instability.*, In 2020 IEEE symposium on security and privacy (SP) (pp. 910-927). IEEE.

[9] Chainspect (2023): *Transactions Per Second (TPS) in top blockchains*, Retrieved November 6, 2024, from `https://medium.com/@chainspect_app/transactions-per-second-tps-in-top-blockchains-001d430dac2b`.

[10] Wikipedia (2024): *Blockchain*, Retrieved November 5, 2024, from `https://en.wikipedia.org/wiki/Blockchain`.

[11] GeeksForGeeks (2024) *What is double spending in blockchain*, Retrieved November 10, 2024, from `https://www.geeksforgeeks.org/what-is-double-spending-in-blockchain/`.

[12] Roopika, J. (2020, October): *Blockchain Technology: History, Concepts, and Applications*, (pp. 645-646).

[13] Crypto Valley Journal (2024, July) *The Blockchain Trilemma*, Retrieved November 6, 2024, from https://cryptovalleyjournal.com/education/basics/the-blockchain-trilemma/.

[14] Investopedia (2024, May) *High-Frequency Trading (HFT): What It Is, How It Works, and Example*, Retrieved November 10, 2024, from https://www.investopedia.com/terms/h/high-frequency-trading.asp.

[15] Coindesk (2021, June) *DeFi Is the Next Frontier of High-Frequency Trading*, Retrieved November 10, 2024, from https://www.coindesk.com/markets/2021/06/23/defi-is-the-next-frontier-of-high-frequency-trading/.

[16] Etherscan (2024, November) *jaredfromsubway.eth Ethereum address*, Retrieved November 10, 2024, from https://etherscan.io/address/0xae2fc483527b8ef99eb5d9b44875f005ba1fae13#analytics.

[17] Etherscan - block 17178637 (2022) *Transactions of Ethereum block number: 17178637*, Retrieved November 10, 2024, from https://etherscan.io/txs?block=17178637&ps=100&p=2.

[18] Eigenphi.io (2024) *Eigenphi.io*, Retrieved November 11, 2024, from https://eigenphi.io/.

[19] Eigenphi.io (2024) *Eigenphi.io sandwich attack victim attack 0x3e28b1d60a47eD10Fa1025d35d772589d6698C0b*, Retrieved November 14, 2024, from https://eigenphi.io/mev/ethereum/sandwich/victim/0x3e28b1d60a47eD10Fa1025d35d772589d6698C0b.

[20] Eigenphi.io (2024) *Eigenphi.io sandwich attacker 0x1f2F10D1C40777AE1Da742455c65828FF36Df387*, Retrieved November 14, 2024, from https://eigenphi.io/mev/ethereum/sandwich/attacker/0x1f2F10D1C40777AE1Da742455c65828FF36Df387.

[21] Wikipedia (2024) *Ethereum Classic*, Retrieved November 14, 2024, from https://en.wikipedia.org/wiki/Ethereum_Classic.

[22] Academy.bit2me.com (2019) *What is a Finney Hack or Finney Attack?*, Retrieved November 14, 2024, from https://academy.bit2me.com/en/que-es-un-hackeo-finney-ataque-finney/.

Chapter 2

Cash-and-Carry Arbitrage for Crypto

Md Rezuanul Haque

This study explores the risks and return of Cash-and-Carry trading strategy in cryptocurrency markets. As cryptocurrency markets grow, they offer new opportunities for trading, including Cash-and-Carry strategy, where traders buy an asset in the spot market and sell it in the futures market. This research looks at the main factors that affect the returns of this strategy, such as investor behavior, margin requirements, convenience yield, and limits on arbitrage. The findings provide useful insights for investors and their risk management concerns. Future research could look into other trading strategies and evaluate their risk-return profiles within a similar framework.

Contents

Contents

2.1 Introduction and Problem Statement

Cryptocurrency markets have grown rapidly over the last decade, attracting a wide range of investors and leading to the creation of various financial products. Among these, derivative trading has become essential, including perpetual and traditional futures contracts. In cryptocurrency markets, perpetual contracts allow traders to hold positions for as long as they want, with no expiration date [17]. This feature has attracted both individual and institutional investors, making the crypto derivatives market an important part of the financial system.

One of the main strategies in cryptocurrency markets is the Cash-and-Carry trade. This strategy involves buying a cryptocurrency in the spot market while selling a futures contract on the same asset. The goal is to profit from the difference, or "carry," between the futures price and the spot price. In traditional financial markets, Cash-and-Carry trading is often a relatively low-risk way to make a profit because liquidity and arbitrage (trading to take advantage of price differences) keep prices closely aligned [18]. However, cryptocurrency markets are very different from traditional markets, particularly in terms of liquidity, arbitrage constraints, and volatility. Crypto markets are usually more volatile, with big price swings that create unique challenges for traders. Additionally, crypto markets can have limited liquidity, and exchanges often have different margin requirements, which can make it harder for larger investors to use this strategy.

2.1.1 Significance of Study

Understanding cryptocurrency markets, Cash-and-Carry strategy is important for speculative and risk management purposes. For speculative traders, crypto carry trades offer high returns that are not often seen in other asset classes, making them very attractive. For those looking to manage risks, Cash-and-Carry trading can help protect against sudden price changes in cryptocurrencies, which are known for their volatility [9]. Cryptocurrencies also have unique characteristics, like decentralization and limited regulation, which make the market less complex and interesting. Unlike traditional assets, cryptocurrencies operate on decentralized networks, meaning economic or political factors less influence them. Additionally, crypto carry trades often present higher carry opportunities than in traditional markets, which could be due to the speculative nature of crypto, its evolving regulations, and the relatively low presence of traditional financial institutions in the market.

2.1.2 Objectives

The objectives of this study are as follows:

1. **Evaluate the Viability of Cash-and-Carry in Crypto Markets:** The first objective is to see if the Cash-and-Carry strategy is viable in cryptocurrency markets. This includes analyzing past returns and risks related to this strategy and identifying when it has been most and least profitable.

2. **Explore Factors Driving Returns in Crypto Cash-and-Carry:** The second objective is to explore the main factors that affect the returns in crypto Cash-and-Carry strategies. This includes looking at the effects of investor sentiment, market volatility, and arbitrage constraints, all of which influence the success of carry trades in crypto markets.

3. **Compare with Traditional Financial Markets:** The third objective is to compare Cash-and-Carry in crypto with Cash-and-Carry in traditional financial markets.

By examining the differences in structure, volatility, and investor behavior, this study aims to understand what makes Cash-and-Carry in crypto both unique and potentially more profitable.

2.1.3 Problem Statement

Cash-and-Carry Trading strategy is used by many practitioners to generate high returns in the cryptocurrency markets. This is while the root cause for the high yields and their corresponding risks are not highlighted. This work aims to elaborate on the risks and return profile of this trading strategy in the framework that were introduced by Abtahi, A., and Abtahi, R., in their working paper titled *"A Taxonomy of Inefficiencies and Arbitrage Opportunities in Cryptocurrency Markets"* [23].

2.2 Basic Definitions (for people without Finance and Economics background)

2.2.1 Arbitrage

Arbitrage is a financial strategy that involves simultaneously buying and selling an asset in different markets to profit from price differences [19]. The key principle of arbitrage is to take advantage of inefficiencies in pricing, ensuring a risk-free profit as the trader capitalizes on the price gap. Arbitrage typically occurs when an asset, such as stocks, commodities, or cryptocurrencies, is priced differently across two or more markets.

For example, if a cryptocurrency like Bitcoin is priced lower on one exchange and higher on another, a trader can buy Bitcoin on the cheaper exchange and sell it on the more expensive one, locking in the price difference as profit. This practice plays a vital role in financial markets by helping to align prices across markets, increasing efficiency and reducing discrepancies over time. However, successful arbitrage often requires speed, access to multiple markets, and the ability to manage transaction costs and liquidity risks.

In cryptocurrency markets, arbitrage opportunities are more frequent due to higher volatility, fragmented markets, and varying liquidity levels.

2.2.2 Future Contract

A futures contract is a legal agreement between two parties to buy or sell a specific asset at a predetermined price on a specified future date [22][15][20]. These contracts are standardized and traded on exchanges, ensuring transparency and reducing counterparty risk. Futures are commonly used for commodities (like oil or gold), financial instruments (such as stocks or bonds), and even cryptocurrencies [26] [18] .

2.2.2.1 Key Features:

1. **Standardization:** Futures contracts specify the quantity, quality, and delivery terms of the asset, making them uniform and easier to trade.

2. **Margin Requirement:** Buyers and sellers are required to deposit an initial margin (collateral) and maintain a margin balance as the market price fluctuates.

3. **Leverage:** Futures allow traders to control large positions with a small initial investment, amplifying potential gains or losses.

4. **Settlement:** Futures contracts can be settled in two ways:
    - *Physical Delivery:* The actual asset is delivered upon contract expiry.
    - *Cash Settlement:* The difference between the agreed price and the market price is paid in cash.

2.2.2.2 Purposes of Futures Contracts:

1. **Hedging:** Businesses use futures to lock in prices and reduce risks related to price fluctuations. For example, farmers might use futures to secure a fixed price for crops.
2. **Speculation:** Traders use futures to bet on the price direction of an asset to earn a profit, without intending to take delivery of the actual asset.
3. **Price Discovery:** Futures markets provide information about expected future prices, helping market participants make informed decisions.

In cryptocurrency markets, futures contracts are widely used for assets like Bitcoin and Ethereum. These contracts help manage the high volatility of crypto assets, allowing traders to hedge risks or speculate on price movements efficiently.

2.2.3 Cash-and-Carry Strategy

Cash-and-Carry Arbitrage is a trading strategy that involves buying an asset in the spot market (the "cash" component) while simultaneously selling a futures contract on the same asset (the "carry" component) [22]. The goal is to lock in a risk-free profit by taking advantage of a price difference between the spot price and the futures price, known as the "futures basis" [21].

2.2.3.1 How it Works:

1. **Spot Purchase:** The trader buys the asset in the spot market at the current market price.
2. **Futures Sale:** At the same time, the trader sells a futures contract for the same asset at a higher price.
3. **Holding the Asset:** The trader holds the asset until the futures contract matures.
4. **Delivery:** On the maturity date, the trader delivers the asset (purchased earlier) to fulfill the futures contract, locking in the profit.

2.2.3.2 Profit Mechanism:

The profit in a Cash-and-Carry arbitrage strategy arises if the futures price is higher than the spot price plus the carrying cost of holding the asset until the contract's maturity. Carrying costs include storage, insurance, financing, and other expenses incurred while holding the asset.

$$\text{Profit} = \text{Futures Price} - (\text{Spot Price} + \text{Carrying Costs})[25]$$

For example, consider a gold arbitrage scenario. A trader buys 1 ounce of gold in the spot market for \$1,800 and simultaneously sells a futures contract for the same ounce at \$1,850. Carrying costs for holding the gold, such as storage and insurance, amount to \$30. The profit calculation would be:

$$\text{Profit} = \text{Futures Price} - (\text{Spot Price} + \text{Carrying Costs}) = 1850 - (1800 + 30) = 20$$

This mechanism illustrates how carrying costs like storage and insurance impact the overall profit in traditional commodity markets. In contrast, carrying a cryptocurrency does not incur physical storage or insurance costs, which simplifies the strategy and often enhances profitability in crypto markets. The absence of these carrying costs in cryptocurrencies makes cash-and-carry arbitrage particularly appealing, especially when futures prices are significantly higher than spot prices.

2.3 Literature Review

2.3.1 Crypto Market Characteristics

Cryptocurrencies have become a major asset class in the past decade, gaining attention for their high returns and distinctive risk profiles. Liu et al.,[12] studied the risk-return profile of cryptocurrencies like Bitcoin, Ripple, and Ethereum and found that these assets are not significantly influenced by traditional financial market factors. Instead, cryptocurrency returns are driven by unique market factors such as momentum and investor attention. The authors showed that search volume and social media activity are strongly linked to short-term returns, suggesting that retail investors play a significant role in price movements. Additionally, the study highlighted the high volatility and positive skewness in cryptocurrency returns, which makes them attractive to risk-seeking investors but also exposes them to frequent market downturns.

In another study, Fan et al.[10] analyzed the carry trade in cryptocurrency markets and found that the cross-sectional strategy, which involves going long on high-interest-rate cryptocurrencies and shorting low-interest-rate ones, can yield annualized returns of 43.4%. This study highlighted the unique risk factors in the cryptocurrency market, which differ from those in traditional assets. Fan et al. also noted that cryptocurrency carry returns are influenced by market volatility and equity market risk. For instance, high-interest-rate cryptocurrencies are more likely to experience losses when equity markets are volatile, leading to a higher risk premium in the crypto market. These findings underscore the unique behaviors and risks present in the cryptocurrency market, which can create both high returns and high risk.

2.3.2 Cash-and-Carry in Cryptocurrency

The Cash-and-Carry strategy in cryptocurrency is increasingly popular but comes with significant challenges. Schmeling et al.[9] examined the dynamics of crypto carry trades and found that the difference between futures and spot prices—known as the futures basis—is often large, with returns reaching up to 60% annually. However, this high return potential is accompanied by substantial volatility. The study identified two main factors driving these dynamics: retail investor behavior and the limited presence of cash-and-carry arbitrage capital in the market. In periods of high market activity, retail traders tend to take long positions in futures, increasing the demand and driving up futures prices. This high demand for futures also makes it more difficult for cash-and-carry arbitrageurs to profit consistently, especially during periods of margin calls and liquidations that arise in volatile market conditions. Thus, while the carry strategy offers high returns, it also exposes traders to increased risk, especially during market drawdowns.

He et al.[11] explored the fundamentals of perpetual futures and basis fluctuations in cryptocurrency markets. Perpetual futures differ from standard futures contracts as they do not have an expiry date, allowing investors to hold positions indefinitely. He et al. found that perpetual futures prices often deviate from spot prices due to convenience yield—an extra value investors place on futures contracts for benefits like high leverage or regulatory ease. The study shows that convenience yields in crypto futures are higher than in traditional markets, largely due to high demand from speculative traders and the challenges that cash-and-carry arbitrageurs face in these markets. Additionally, the authors highlighted the impact of funding rate adjustments on perpetual futures prices, which further adds complexity to crypto carry trades.

2.3.3 Comparison with Traditional Markets

The Cash-and-Carry strategy is widely used in traditional markets, such as fiat currencies and commodities, but the dynamics in crypto markets differ in several important ways. In traditional markets, Cash-and-Carry typically involves buying a spot asset and selling a futures contract, with minimal risk because futures prices tend to closely track spot prices due to high liquidity and active arbitrage. For example, the futures basis in fiat currency markets, which is largely driven by interest rate differentials, remains relatively stable. However, in cryptocurrency markets, Schmeling et al.[9] noted that the basis can be volatile due to factors like margin requirements, regulatory restrictions, and high leverage limits, which are not as common in traditional markets. Additionally, while commodity markets experience some volatility, it is generally less extreme than in crypto markets, where futures prices can deviate significantly from spot prices due to retail speculation and arbitrage limitations [11].

The differences highlight the unique challenges and opportunities for Cash-and-Carry strategies in crypto markets compared to traditional financial markets. Cryptocurrency markets offer potentially higher returns but come with increased exposure to volatility and liquidity constraints, which make arbitrage more challenging. By studying these unique characteristics, researchers can better understand the broader implications for carry trades in cryptocurrency and other high-risk markets.

2.4 Cash-and-Carry Arbitrage Attributes

In this section we will dive into the Cash-and-Carry trading strategy attributes in relation to cryptocurrency arbitrage, based on the author's suggested framework in [23].

1. **Ethical/Unethical**

   Cash-and-Carry is typically considered an ethical trading practice because it seeks to exploit price inefficiencies without disrupting the market. Ethical arbitrage aligns with market principles and contributes to price efficiency across platforms. However, unethical practices such as wash trading, which artificially inflates trading volume to manipulate market prices, or frontrunning, where traders use prior knowledge of pending orders to exploit market positions, can blur ethical boundaries.

2. **Directional (Delta-Neutrality)**

   Directional vs. Delta-Neutral Arbitrage strategies can be categorized based on their reliance on market price movements. Delta-neutral strategies, such as Cash-and-Carry arbitrage, do not depend on directional market trends, as profits are derived from price discrepancies between the spot and futures markets. These strategies are inherently less risky and more stable than directional approaches. In contrast,

directional arbitrage strategies, like statistical arbitrage, leverage predictive models or market trends to forecast price movements [6; 7]. Delta-neutrality is particularly advantageous in volatile markets like cryptocurrencies, where price fluctuations can be sudden and extreme.

3. **Capital Intensive**

   Many cryptocurrency arbitrage strategies are capital-intensive due to high transaction costs, withdrawal fees, and the need to maintain sufficient liquidity across multiple exchanges. For example, cross-border arbitrage, such as exploiting the "Kimchi Premium" in South Korea, often requires significant capital to cover regulatory and logistical constraints [5]. Similarly, Cash-and-Carry strategies necessitate substantial upfront investment to secure positions in both spot and futures markets. Additionally, the lack of cross-margining in many cryptocurrency exchanges amplifies the capital required for effective risk management [23].

4. **Historical Returns**

   Cash-and-Carry Arbitrage opportunities in cryptocurrency markets have historically provided higher returns compared to traditional markets, especially during periods of high market inefficiency. For instance, during 2017's crypto boom, price discrepancies across exchanges frequently exceeded 10%, offering substantial profit margins [9; 10]. However, as the market matures, increased competition, regulatory oversight, and technological advancements are narrowing these opportunities. The use of automated trading bots and sophisticated algorithms by institutional players has further reduced the frequency and profitability of arbitrage opportunities.

   When it comes to the **Cash-and-Carry trading strategy**, the returns are largely derived from the futures basis, which is the difference between the futures price and the spot price of a cryptocurrency. This basis tends to widen during periods of speculative demand in the futures market, creating profitable opportunities for arbitrageurs. For instance, Bitcoin-based Cash-and-Carry trades have historically delivered annualized returns ranging between 8% and 15%, while Ethereum has averaged slightly lower returns of 6% to 12% during certain periods [9].

   However, these returns are not without risk. The profitability of Cash-and-Carry trades can be influenced by market conditions, including sudden price volatility, changes in margin requirements, and transaction costs. Furthermore, the lack of cross-margining on many cryptocurrency exchanges adds to the capital intensity of these trades, which can reduce their overall profitability. Despite these challenges, Cash-and-Carry remains a favored strategy for arbitrageurs due to its relatively delta-neutral nature, making it less reliant on directional market movements.

5. **On-Chain/Off-Chain**

   The Cash-and-Carry trading strategy can be implemented both on-chain (through decentralized exchanges or DEXs) and off-chain (through centralized exchanges or CEXs). However, it has primarily been used in off-chain methods on CEXs due to being more straightforward than DEXs. Typically, On-chain methods require technical expertise and face smart contract risks, whereas off-chain methods demand financial acumen and carry counterparty risks. Here arbitrage trading strategies are classified based on the execution method:

   - **On-Chain Arbitrage**: This involves transactions executed directly on the blockchain, often leveraging decentralized exchanges and DeFi protocols. Examples include flash loan arbitrage, which uses non-collateralized loans to exploit price inefficiencies within a single transaction [1; 2], and bridge arbitrage, which transfers

tokens between blockchains to benefit from cross-chain price differences [3]. Wang et al. [2] further explore flash loan applications, highlighting the mechanics of executing arbitrage within a single transaction cycle.

- **Off-Chain Arbitrage**: This relies on centralized exchanges and external platforms, such as statistical arbitrage using machine learning models to identify inefficiencies [7], or Cash-and-Carry strategies involving spot and futures markets [9].

6. **Root Cause**

   The study done by BIS, explores the root cause of the persisting high returns of the Cash-and-Carry Trade for the crypto markets [9]. It outlines that the so-called Trend-Chasing Investors tend to take levergaed long positions in the futures market, specially during the periodic bullruns. This behavior pushes the future's prices high. In a more established asset class, the larger investors (specially the yield chasing investors) will offset this demand by providing capital and investing in Cash-and-Carry strategies to farm additional interests. This is while in the crypto market the larger institutions cannot fully offset the demand pressure from the trend chasing investors (usually retail investors) due to key frictions, such as the lack of cross-margning between spot and futures positions especially for the institutionalized exchanges like CME. This makes exploiting the arbitrage risky for the larger institutions, as widening the spreads can trigger margin calls, forcing liquidations before prices converge. Consequently, carry traders are cautious, and the scarcity of arbitrage capital allows the basis to remain elevated.

7. **Durability**

   Arbitrage opportunities in cryptocurrency markets are often short-lived, as market forces quickly align prices through arbitrage activities. However, structural inefficiencies, such as those caused by regulatory gaps or limited institutional participation, can extend the duration of certain opportunities. For example, Cash-and-Carry arbitrage profits have remained elevated for extended periods due to constraints like high capital requirements and limited arbitrage capital from institutional investors [9].

8. **Risks**

   Cryptocurrency arbitrage, despite its potential for profit, carries significant risks:

   - **Price Volatility:** Sudden price fluctuations can erode profits during the execution of arbitrage trades. For instance, Schmeling et al. [9] found that futures basis deviations during volatile periods, such as March 2020, exceeded 5%, creating challenges for traders despite lucrative opportunities.
   - **Network Congestion:** Delays in transaction processing, particularly in on-chain methods, can lead to missed opportunities. Flash loans, as discussed by Qin et al. [1], and Wang et al. [2], exacerbate network congestion during periods of high arbitrage activity, further complicating timely execution.
   - **High Fees:** Transaction and withdrawal fees can significantly reduce arbitrage margins. Kannengießer et al. [3] emphasize how cross-chain technology introduces additional costs that limit profitability in bridge arbitrage scenarios.
   - **Counterparty Risk:** Exchange failures or defaults can result in loss of funds, especially in off-chain methods reliant on centralized platforms. Kim et al. [17] highlight the role of decentralized perpetual contracts in reducing counterparty risks, but their adoption remains limited.

- **Regulatory Uncertainty:** Arbitrage across jurisdictions is subject to varying legal and tax implications, which can affect both feasibility and profitability. Regulatory complexities, as noted by Conlon et al. [16], can significantly influence the arbitrage landscape, particularly for cross-border trades.

By addressing these attributes comprehensively, this taxonomy provides a framework for understanding the dynamics of cryptocurrency cash-and-carry arbitrage and serves as a guide for practitioners and researchers aiming to optimize strategies in this complex and evolving market.

2.5 Results and Discussion

2.5.1 Profitability and Viability of Crypto Cash-and-Carry

The analysis revealed that Cash-and-Carry strategies in cryptocurrency markets are viable and profitable, though they come with unique challenges. Bitcoin-based Cash-and-Carry trades consistently delivered annualized returns ranging from 8% to 15%, while Ethereum averaged slightly lower returns at 6% to 12% [9]. These returns are significantly higher than those observed in traditional fiat currency or commodity carry trades, which typically yield 2% to 5%, as highlighted by Fan et al.[10]. However, the elevated returns in crypto markets are accompanied by considerable risks, including sharp drawdowns during periods of extreme volatility. For instance, Schmeling et al.[9] observed that futures basis deviations during March 2020 exceeded 5%, creating short-lived but highly lucrative arbitrage windows.

2.5.2 Key Drivers of Crypto Cash-and-Carry Returns

The success of Cash-and-Carry trades is closely tied to market conditions and specific factors. Convenience yield emerged as a critical driver of futures premiums. Investors value the leverage and reduced custody risks offered by futures, which are reflected in high funding rates for perpetual futures, averaging 0.01% per hour, as noted by He et al.[11]. Market volatility, both in the cryptocurrency market and in traditional financial markets, significantly influences futures basis.
According to a study by Conlon et al.[16], there is a noticeable correlation between the volatility of the stock market and the behavior of Bitcoin futures. Specifically, when the stock market experiences high volatility, as indicated by the VIX—a measure known as the ”fear gauge”—there tends to be a higher basis level in Bitcoin futures. This increase in the basis level can be interpreted as a sign of heightened risk aversion among investors, as well as an increase in speculative demand for Bitcoin. Essentially, during times of uncertainty and fear in traditional markets, investors might turn to Bitcoin futures either as a hedge against their other investments or as a speculative opportunity to profit from market fluctuations.

Investor sentiment plays a pivotal role in driving the profitability of Cash-and-Carry strategies. Positive sentiment, as measured by increased search trends and heightened social media activity, tends to widen the futures basis, creating more lucrative opportunities. Liu and Tsyvinski[13] found that a 10% increase in Google search interest for Bitcoin correlates with a 5% increase in basis spreads. However, these sentiment-driven opportunities are highly sensitive to liquidity constraints. Lower-liquidity exchanges like Bitfinex often exhibit more pronounced and persistent deviations compared to high-liquidity platforms such as Binance, where cash-and-carry arbitrage opportunities close within minutes, as shown by Shynkevich[24].

2.5.3 Comparative Insights from Traditional Financial Markets

Comparatively, cryptocurrency Cash-and-Carry strategies differ significantly from their counterparts in traditional financial markets. While traditional markets benefit from institutional participation and centralized liquidity, cryptocurrency markets are fragmented and heavily influenced by retail speculation. This fragmentation amplifies inefficiencies, creating cash-and-carry arbitrage opportunities that are less common in traditional markets. Krueger et al.[14] highlighted that events like Ethereum's transition to Proof-of-Stake create temporary basis surges, reflecting speculative demand and insider trading in crypto markets. In contrast, traditional markets exhibit smaller, more stable basis deviations due to tighter regulatory oversight and higher liquidity.

2.5.4 Risk-Reward Dynamics in Cryptocurrency Markets

Despite these challenges, the study found that Cash-and-Carry strategies in cryptocurrency markets offer a compelling risk-reward profile for experienced investors. While traditional markets provide stability and predictability, the higher returns and dynamic nature of cryptocurrency markets make them uniquely attractive for cash-and-carry arbitrage strategies. The findings underscore the importance of understanding market-specific drivers and the impact of broader macroeconomic factors on futures pricing.

2.6 Summary and Conclusion

- **Summary**: This paper investigates the viability and profitability of Cash-and-Carry arbitrage in cryptocurrency markets, focusing on major assets like Bitcoin and Ethereum. It examines the eight key attributes of arbitrage, including delta-neutrality, historical returns, and associated risks, while highlighting how cryptocurrency markets differ from traditional financial markets in terms of volatility, liquidity, and regulatory structures. The study reveals that Cash-and-Carry arbitrage in cryptocurrencies offers higher returns compared to traditional markets, driven by speculative investor behavior and market inefficiencies. However, these opportunities are accompanied by significant risks such as price volatility, liquidity constraints, and regulatory uncertainty. By providing a comprehensive analysis, the paper bridges the knowledge gap in understanding the unique dynamics of cryptocurrency arbitrage and suggests directions for future research, including stablecoin-based strategies and the impact of cross-margining mechanisms.

- **Conclusion**: Cash-and-Carry arbitrage in cryptocurrency markets presents a compelling opportunity for investors, offering significantly higher returns than traditional markets due to speculative demand and structural inefficiencies. However, these opportunities are inherently risky, with challenges such as extreme market volatility, fragmented liquidity, and regulatory ambiguity. The study emphasizes the importance of tailored risk management approaches to navigate these risks effectively.

Bibliography

[1] K. Qin, L. Zhou, B. Livshits, and A. Gervais, "Attacking the DeFi ecosystem with flash loans for fun and profit," in *International Conference on Financial Cryptography and Data Security*, Springer, Berlin, Heidelberg, pp. 3–32, 2021.

[2] D. Wang, S. Wu, Z. Lin, L. Wu, X. Yuan, Y. Zhou, ... and K. Ren, "Towards a first step to understand flash loan and its applications in DeFi ecosystem," in *Proceedings of the Ninth International Workshop on Security in Blockchain and Cloud Computing*, pp. 23–28, 2021.

[3] N. Kannengießer, M. Pfister, M. Greulich, S. Lins, and A. Sunyaev, "Bridges between islands: Cross-chain technology for distributed ledger technology," 2020.

[4] Binance, "Frontrunners and MEV explained: How to beat the bots," *Binance*, November 21, 2024. Retrieved from Binance. `https://www.binance.com/en/square/post/188297`.

[5] Investopedia, "Kimchi premium," *Investopedia*, May 11, 2024. Retrieved from Investopedia.

[6] dYdX, "Statistical arbitrage," *dYdX*, November 21, 2024. Retrieved from dYdX. `https://dydx.exchange/crypto-learning/statistical-arbitrage`.

[7] T. G. Fischer, C. Krauss, and A. Deinert, "Statistical arbitrage in cryptocurrency markets," *Journal of Risk and Financial Management*, vol. 12, no. 1, pp. 31, 2019.

[8] K. Qin, L. Zhou, and A. Gervais, "Quantifying blockchain extractable value: How dark is the forest?" in *2022 IEEE Symposium on Security and Privacy (SP)*, IEEE, pp. 198–214, 2022.

[9] M. Schmeling, A. Schrimpf, and K. Todorov, "Crypto carry," *Available at SSRN 4268371*, 2023.

[10] Z. Fan, F. Jiao, L. Lu, and X. Tong, "The risk and return of cryptocurrency carry trade," *Available at SSRN*, 2023.

[11] Chengying He, Yong Li, Tianqi Wang, and Salman Ali Shah, "Is cryptocurrency a hedging tool during economic policy uncertainty? An empirical investigation," *Humanities and Social Sciences Communications*, vol. 11, no. 1, pp. 1–10, 2024, Palgrave.

[12] Y. Liu, A. Tsyvinski, and X. Wu, "Common risk factors in cryptocurrency," *The Journal of Finance*, vol. 77, no. 2, pp. 1133–1177, 2022, Wiley Online Library.

[13] Y. Liu and A. Tsyvinski, "Risks and returns of cryptocurrency," *The Review of Financial Studies*, vol. 34, no. 6, pp. 2689–2727, 2021, Oxford University Press.

[14] S. Krueger, M. Müller, A. Betzer, and C. Rokitta, "Event-driven strategies in crypto assets," *Available at SSRN 3363589*, 2019.

[15] S. He, A. Manela, O. Ross, and V. von Wachter, "Fundamentals of perpetual futures," *arXiv preprint arXiv:2212.06888*, 2022.

[16] T. Conlon, S. Corbet, and L. Oxley, "Investor sentiment, unexpected inflation, and Bitcoin basis risk," *Journal of Futures Markets*, vol. 44, no. 11, pp. 1807–1831, 2024, Wiley Online Library.

[17] H. Kim, H. S. Kim, and Y. Park, "Perpetual contract NFT as collateral for DeFi composability," *IEEE Access*, vol. 10, pp. 126802–126814, 2022, IEEE.

[18] I. Bouchouev, "Virtual barrels," *Springer Texts in Business and Economics*, 2023, Springer.

[19] J. M. Haykov, "Arbitrage money," *Journal of Critical Realism in Socio-Economics (JOCRISE)*, vol. 2, no. 2, pp. 219–241, 2024.

[20] V. Lucic and A. Sepp, "Valuation and hedging of cryptocurrency inverse options," *Quantitative Finance*, vol. 24, no. 7, pp. 851–869, 2024, Taylor & Francis.

[21] S. K. Parameswaran, *Derivatives Theory and Practice: An Emerging Markets Perspective*, Walter de Gruyter GmbH & Co KG, 2024.

[22] E. Chen, M. Ma, and Z. Nie, "Perpetual future contracts in centralized and decentralized exchanges: Mechanism and traders' behavior," *Electronic Markets*, vol. 34, no. 1, pp. 35, 2024, Springer.

[23] R. Abtahi and S. A. Abtahi, "A taxonomy of inefficiencies and arbitrage opportunities in cryptocurrency markets," University of Zurich, Department of Informatics, Working Paper, 2024.

[24] Andrei Shynkevich, 'Law of one price and return on Arbitrage Trading: Bitcoin vs. Ethereum," *Journal of Economics and Finance*, vol. 47, no. 3, pp. 763–792, 2023, Springer.

[25] J. Hull, *Options, Futures, and Other Derivatives*, 11th ed., Pearson, 2021.

[26] SNC van Rij, 'An Empirical Analysis of Cost-of-Carry and Quarterly Futures Prices in the Cryptocurrency Market," *Available at SSRN*, 2023.

Chapter 3

The Role of Decentralized Identities in Central Bank Digital Currency (CBDC)

Raphael Duka, 18-107-904

Central Bank Digital Currencies (CBDCs) will transform the global financial ecosystem, offering expanded financial inclusion, optimized efficiency, and a modernized monetary framework. Decentralized identity frameworks can help address privacy, security, and public trust challenges. They align with self-sovereignty and privacy principles, giving users control over their data, which helps CBDC adoption. The study identifies models and initiatives that integrate decentralized identities into CBDC ecosystems. A review of operational CBDCs highlights the need for an optimal environment for adoption. The paper discusses policy recommendations, technical solutions, and public engagement strategies to enable seamless integration and promote public trust. The paper highlights global interest in decentralized identity frameworks, with notable advancements in this field by the European Union. Innovative policies and frameworks can establish a secure, inclusive, and user-centric digital financial ecosystem. This paper concludes that such advancements could set a global benchmark for the coexistence of financial inclusion and user autonomy in the digital economy.

Contents

3.1 Introduction

3.1.1 Context of CBDC Development and The Role of Decentralized Identities

As financial transactions have increasingly moved into wireless and cloud-based systems, the global shift away from cash to digital payments has accelerated [1]. This trend was significantly intensified by the COVID-19 pandemic, which created a growth in demand for contactless payments to reduce physical interactions [2; 3]. The pandemic also contributed to the rising popularity of digital assets, such as cryptocurrencies, as consumers and businesses sought flexible, digital-first payment methods [4; 5]. This transition has not only shown the importance of digital payment infrastructure but has also pushed central banks worldwide to consider issuing their own electronic currency.
Central Bank Digital Currencies (CBDCs) aim to combine the accessibility and security of central bank money with the efficiency of digital payments, offering a state-backed digital alternative in the evolving financial landscape [6]. Globally, over 100 countries, representing 95% of global GDP, are exploring CBDC projects [7], with four CBDCs already having been successfully introduced as legal tender [8]. Central banks, governments and policy makers around the world are interested in the introduction of CBDCs as they promise improved efficiency, increased financial inclusion, and enhanced payment security [7]. As the CBDC projects progress, however, several challenges have emerged, particularly regarding privacy and data security. Since CBDCs often require transaction traceability for regulatory compliance, there is a risk of creating detailed digital records of users' financial behaviors, which could be accessed by authorities, potentially violating personal privacy. This privacy concern is where decentralized identities can play a significant role. Decentralized identities allow individuals to manage their digital identities directly, granting them more control over their personal information and limiting the amount of data exposed during transactions [9]. By using decentralized identity frameworks, CBDCs can make sure that users authenticate their identities securely while minimizing personal data exposure. The decentralized identity framework, uses cryptographic methods to keep users' data private and secure. A successful integration of such a system into CBDC will promote the implementation and adoption of Central Bank Digital Currencies, making digital transactions more secure and widely accessible.

3.1.2 Problem Statement and Structure

This report will explore the complexities of integrating decentralized identities into CBDC systems, beginning with an overview of the technical foundations of CBDCs and decentralized identities. The technical framework for CBDCs involves an infrastructure that must guarantee security, reliability, and scalability to support large-scale financial transactions. On the other hand, decentralized identity systems rely on cryptographic techniques, which allow individuals to control their digital identities in a secure and private manner. By examining these elements, this report will establish a foundation for understanding the interaction between CBDCs and decentralized identities.
In addition to the technical foundations, I will also address the limitations and potential barriers to integrating decentralized identities within CBDC systems. These include significant regulatory obstacles, as decentralized identity frameworks present unique challenges for existing regulatory standards that are designed around more traditional, centralized identification systems. The report will explore how these issues could impact the practical implementation of decentralized identities in CBDCs, as well as how they might shape user adoption and public trust.

Following this, this report will examine current CBDC projects to understand how decentralized identity solutions are being approached in real-world applications. By studying existing projects from various jurisdictions, the analysis will show the potential benefits and limitations of decentralized identities in these environments. These examples will highlight specific advantages, such as user privacy and reduced data exposure, as well as limitations, including challenges in scaling identity solutions and achieving interoperability across different systems.
Finally, this report will propose strategies for overcoming the barriers identified, offering potential ways for integrating decentralized identities into CBDC systems. This section will outline practical solutions aimed at addressing regulatory and technical challenges, along with policy suggestions to create a more supportive environment for decentralized identity adoption within CBDC systems. These strategies will include recommendations for designing interoperable frameworks, establishing privacy-preserving standards, and advancing collaboration between central banks, regulatory bodies, and technology providers. The findings aim to offer a balanced perspective on both the potential benefits and the challenges of decentralized identities in the landscape of digital currencies. With this examination, this report will seek to contribute to the discussion on how to design digital currency systems that respect individual privacy while ensuring the robustness and trustworthiness required in state-backed financial infrastructures.

3.2 Technical Foundations, Challenges and Limitations

The development of CBDCs reflects a significant shift in the financial landscape, as central banks, governments, and policymakers worldwide recognize the need for a digital form of currency that aligns with modern economic demands. CBDCs present an opportunity to sustain the role of central bank money within an increasingly digital economy. They are designed to promote financial inclusion by providing direct access to digital central bank money and financial services without the need of a bank account [11] , improve payment efficiency, and strengthen resilience within payment systems.
This section begins by examining the frameworks of CBDCs, analyzing their structures, types, and the supporting technologies. A thorough exploration of CBDC frameworks will show the foundational elements and design considerations that central banks must address to fulfill their objectives. Following this, I will investigate decentralized identity frameworks, which offer the promise of greater privacy, user control, and security in digital currency ecosystems. However, integrating decentralized identities within CBDC systems presents big challenges and understanding these barriers is crucial for developing effective solutions to overcome them.

3.2.1 CBDC Frameworks

Central Bank Digital Currencies are designed to keep the stability and public trust associated with traditional central bank money while modernizing it for a digital economy [10]. Some of the key design choices to be considered for issuing CBDCs are identified and discussed in this section.
Retail CBDCs (rCBDC): Retail CBDCs are intended for use by the general public, functioning similarly to cash [1]. They enable individuals and businesses to use a secure, state-backed digital currency for their everyday transactions. In regions where financial inclusion is a high priority, rCBDCs offer particular value by allowing universal access to digital currency without the need for a bank account [11]. These CBDCs are often designed with ease of use, offline capabilities, and enhanced security in mind, ensuring accessibility for all segments of the population.

Wholesale CBDCs (wCBDC): Wholesale CBDCs are tailored for financial institutions, enabling them to carry out high-value interbank transactions with improved efficiency. These CBDCs are primarily focused on enhancing the efficiency and security of the interbank settlement process, particularly for cross-border payments [10]. Wholesale CBDCs frequently use distributed ledger technology (DLT) or blockchain to enable seamless, transparent, and nearly instantaneous transactions, addressing traditional challenges in international banking such as lengthy settlement times and high transaction costs [12].
From a technical standpoint, CBDCs can adopt different models, with two primary structures being account-based and token-based systems.
Account-Based Model: In an account-based CBDC system, individual transactions are recorded by updating the account balances of users, resembling conventional banking systems [14]. Each transaction requires verification of the account holder's identity to prevent double-spending and fraud, thus requiring a robust digital identity scheme. Account-based models are often preferred for rCBDCs in jurisdictions where regulatory oversight and traceability are prioritized [13].
Token-Based Model: A token-based CBDC operates similarly to cash, allowing CBDC tokens to be transferred between parties without being linked to specific accounts [15]. Ownership is verified through the validity of the token rather than the identity of the account holder, which accesses the CBDC based on a password-like digital signature using private-public key cryptography, without requiring any personal identification, making the token-based CBDC systems more privacy-preserving than account-based models [16]. This model is beneficial in scenarios where anonymity is prioritized, but it also demands mechanisms to prevent double-spending and unauthorized use [12].
In practice, many central banks are exploring hybrid models that combine features from both account-based and token-based frameworks, aiming to balance privacy with security. Hybrid models often involve a central ledger managed by the central bank, combined with a distributed ledger for transaction processing, allowing for decentralized verification while maintaining central oversight.
The Technology of rCBDCs: *Auer and Boehme (2020)* [21] provide an architecture of such a hybrid retail CBDC, as can be seen in figure 3.1.

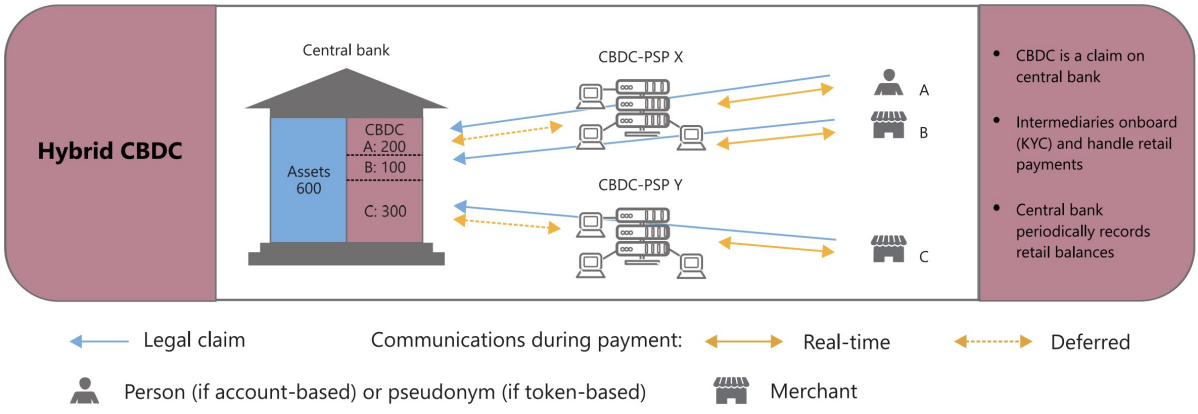


Figure 3.1: Hybrid retail CBDC architecture (Source: *Auer and Boehme (2020)* [21])

As pointed out by *Auer and Boehme (2020)* [21] the architecture of rCBDCs is designed to ensure robustness, efficiency, and inclusivity in digital transactions while safeguarding values such as privacy and security. The underlying technologies for rCBDCs integrate elements of distributed ledger technology, cryptographic mechanisms, and secure communication protocols to deliver a reliable digital payment system. By decentralizing the verification process, DLT reduces reliance on a single point of failure and enhances resilience against cyberattacks.

Cryptographic mechanisms, particularly private-public key cryptography, play an important role in securing rCBDC systems. These techniques ensure the authenticity of transactions and protect against unauthorized access. In token-based models, cryptography is fundamental in preventing double-spending and validating token ownership without requiring intermediary authentication processes. Offline transaction capabilities are another important technological aspect, designed to ensure access and inclusivity. This feature allows transactions to be completed without internet connectivity, which is critical in remote areas or during network outages [21; 13].
Retail CBDCs are also developed to ensure integration with existing payment systems and financial infrastructure. This interoperability allows rCBDCs to function alongside conventional banking systems, enabling individuals and businesses to adopt digital currencies without substantial disruptions to their financial practices [22]. Privacy is a key consideration in the design of retail CBDCs, and advanced privacy-enhancing technologies such as zero-knowledge proofs are often utilized to protect user anonymity while meeting regulatory requirements. These technologies aim to balance the conflicting goals of user privacy and traceability, ensuring compliance with legal standards without compromising individual freedoms.
Scalability and performance optimization are essential for rCBDCs. Achieving scalability often involves optimizing consensus mechanisms in DLT or implementing hybrid systems where critical functions are managed centrally while less sensitive operations are distributed. Cybersecurity is another critical aspect, with advanced frameworks designed to protect against fraud, cyberattacks, and unauthorized access.
Challenges and Limitations: Despite their potential benefits, CBDCs present significant challenges and limitations that require careful consideration. One of the most apparent concerns is privacy, as CBDCs often involve the central handling of sensitive personal and transactional data. To comply with regulatory requirements such as Anti-Money Laundering (AML) and Countering the Financing of Terrorism (CFT), CBDC systems must maintain a level of traceability [23; 9]. However, this conflicts with users' expectations of privacy. The integration of electronic identification (eID) into CBDC wallets, as suggested in some designs, could mitigate risks of fraud [11; 24] but simultaneously heightens concerns about surveillance and data misuse [9].
Security risks are another critical limitation. The centralized nature of CBDCs, combined with their reliance on complex technological infrastructure, makes them vulnerable to cyberattacks. Fraud and identity theft are particular risks in CBDC scenarios, where robust verification mechanisms such as cryptographic solutions are needed [9].
Decentralized indentity systems are emerging as a possible solution to these concerns, enabling a successful adoption of rCBDCs. As highlighted by *Johnson (2024)* [25] and *Adams et al. (2021)* [9], decentralized identity frameworks provide individuals with greater control over their personal data, minimizing dependency on centralized authorities and significantly reducing the risk of surveillance and misuse. By decentralizing identity verification and integrating strong cryptographic protections, decentralized identities not only enhance privacy but also align with the regulatory requirements for traceability and compliance, such as AML and CFT. Incorporating decentralized identity systems into retail CBDC infrastructure offers a robust way to balance security, privacy, and public trust, hence making wider acceptance and usability possible.

3.2.2 Decentralized Identity Frameworks

Adams et al. (2021) [9] propose a comprehensive framework for implementing decentralized identities within the context of CBDCs. Their approach integrates Self-Sovereign Identity (SSI) principles with qualified electronic signature standards, aiming to ensure privacy, regulatory compliance, and security. This framework builds on the European

eIDAS (electronic Identification, authentication, and trust services) regulation [26] and uses cryptographic methods to create a robust system of trust and authentication.
At the core of their proposal is the concept of decentralized identity, where each user is assigned a unique digital identity managed via an eID wallet. This wallet operates on cryptographic principles, utilizing public-private key pairs to ensure secure ownership and usage. The identity provider plays a central role in issuing credentials, which bind the user's real-world identity to their digital counterpart. These credentials include qualified certificates for electronic signatures and are issued in compliance with eIDAS standards. The decentralized identity system separates identity from attributes, ensuring that users can prove specific characteristics (such as age or residency) without revealing unrelated personal information. Attribute certificates, issued by trusted attribute providers, bind additional verifiable credentials to the wallet. For example, a university might issue a credential verifying that an individual is a graduate or give information on other educational qualifications, while a government agency might confirm their age. This approach minimizes data exposure, safeguarding user privacy while allowing regulatory oversight.
The implementation extends to qualified electronic signatures. *Adams et al. (2021)* [9] propose a structure where each transaction or interaction generates a verifiable proof that binds identity and attributes securely. Figure 3.2 illustrates the process of creating a qualified signature within this framework. The process begins with the verifier preparing a challenge, a piece of data to be signed. In a CBDC setting, this could be a merchant, requesting certain characteristics (e.g. age) of a buyer before the transaction can go through. In The buyer's eID wallet combines this challenge with their identity credentials and signs the resulting package using their private key. It is possible to include the person's eID wallet certificate (i.e. their qualified certificate) in a Java Web Signature (JWS) structure, as illustrated in figure 3.2. In order to enable offline verification, the complete certificate chain of the aforementioned certificate must also be incorporated into the JWS. The signed proof hence includes both the verifiable credentials and the complete certificate chain in the JWS, in order to make sure that CBDC transactions can be made in offline settings as well.

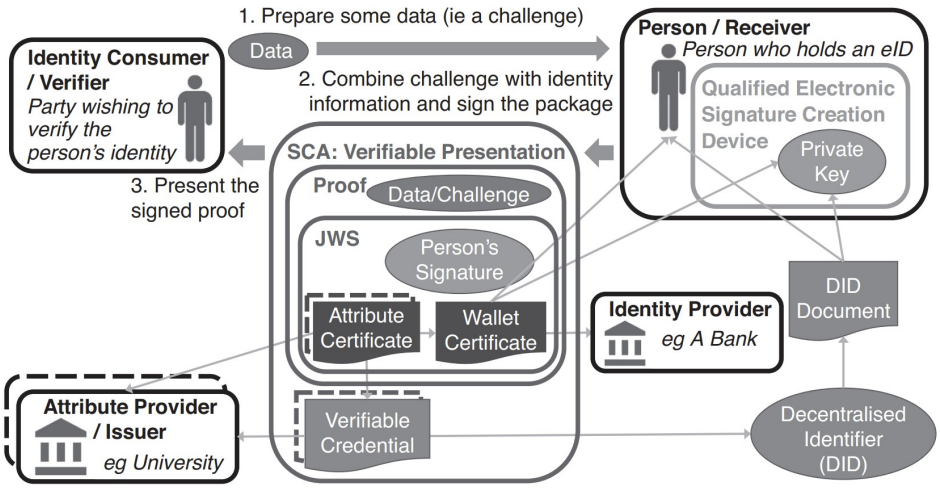


Figure 3.2: eID qualified signature proof in a W3C (SSI) format (Source: *Adams et al. (2021)* [9])

Figure 3.2 is crucial for understanding the interplay between decentralized identities and qualified signatures. It shows how each entity in the system interacts through cryptographically secure data exchange. The user's e-ID wallet credentials, represented by decentralized identifier (DID) documents, link to verifiable credentials issued by trusted providers. These credentials, together with the user's private key stored in a secure element (e.g.,

within the separated secure execution environment through the software installed inside a smartphone [27]), form the foundation for creating and validating digital signatures.
The framework places an emphasis on interoperability, thereby supporting the integration with existing financial systems. Furthermore, the framework addresses offline capabilities, including provisions for local certificate storage and verification. This guarantees the system's resilience even in the absence of a network connectivity.
Adams et al. (2021)'s framework outlines a practical pathway for integrating SSI principles with qualified electronic signatures in CBDCs. By decentralizing identity management while maintaining regulatory compliance, their approach balances privacy, security, and functionality, paving the way for user-centric financial systems.
The integration of decentralized identity frameworks into CBDC represents a promising way for achieving a balance between user privacy, security, and regulatory compliance. The empowerment of individuals through the control of their digital identities can enhance trust and facilitate the broader adoption of CBDCs, as enabled by SSI frameworks. This approach is consistent with global trends toward decentralized identity solutions, as shown by the World Economic Forum's (WEF) examination of decentralized identity principles and their potential to transform digital interactions [28]. Moreover, the International Monetary Fund (IMF) underscores the significance of inclusive strategies for the adoption of central bank digital currencies, emphasizing the necessity for frameworks that address both user needs and regulatory requirements [29]. By adopting SSI principles, central banks can develop CBDC systems that are not only secure and efficient but also respectful of user autonomy and privacy, thereby creating greater public trust and participation in the digital economy.

3.2.3 Barriers to Integration in CBDC Systems

The World Economic Forum's paper, *Reimagining Digital ID* [28], provides a comprehensive analysis of the barriers to successfully integrating decentralized identity frameworks into (r)CBDCs. These challenges, categorized as technical, policy-related, governance, and implementation hurdles, illustrate why decentralized identities have yet to achieve widespread adoption. This section will discuss these barriers, providing a foundation for the discussion in section 3.4 on strategies to effectively overcome these obstacles.
One of the most pressing challenges is technical immaturity. The underlying technologies that support decentralized identity systems, such as zero-knowledge proofs (ZKP) and verifiable credential (VC) standards, remain under development [30] (for an explainer and a literature overview on VC standards see [31]). Frequent updates to standards, like the W3C VC Data Model, complicate the alignment of stakeholders on consistent technical protocols. Without standardized practices, interoperability, critical for seamless data exchange across systems, remains difficult to achieve. A lack of interoperability can lead to vendor lock-in, where individuals are tied to specific providers, undermining the user-centric goals of decentralized identities.
The user experience and accessibility aspects serve to further intensify the technical barriers that already exist. The management of cryptography-based assets, such as private keys, remains a challenge for the typical user [32]. The current lack of intuitive user interfaces and recovery mechanisms in decentralized identity systems may act as deterrents to adoption. Furthermore, the steep learning curve associated with using such systems highlights the necessity of user education and support to build digital literacy. Infrastructure limitations, including scalability issues with distributed ledger technologies that use certain consensus mechanisms, such as proof-of-work, present another challenge [33].
Policy-related barriers include the absence of high-assurance official identification systems in many regions. Without robust identity-binding mechanisms provided by governments, decentralized identity systems cannot achieve their full potential. For example, more than

21 million individuals in the United States (11 % of US citizens), lack official identification, which limits their access to digital identity frameworks [34; 35]. Moreover, regulations in some jurisdictions are not yet aligned to support the reuse of credentials for Know Your Customer (KYC) processes. These gaps in enabling policy frameworks discourage the development and adoption of decentralized identity systems [28].

The absence of effective governance, implementation strategies, and transparent communication represents a significant obstacle to the widespread adoption of decentralized identity systems. It is important that governance frameworks align diverse stakeholders on critical aspects such as roles, responsibilities, and liability. Doing so is essential for ensuring trust and scalability within these systems. In the absence of such frameworks, the creation of a coherent and interoperable ecosystem becomes highly challenging.

A particularly critical challenge is the communication barrier. It is inherently challenging to convey the advantages of decentralized identity systems to the general public, largely due to the intricate technical aspects involved, including verifiable credentials and decentralized identifiers [28]. Without effective communication strategies, the potential advantages of decentralized identity systems, such as enhanced privacy, user control, and data security, remain unclear to the general public. This lack of clarity can discourage adoption, even in contexts where such systems could significantly enhance personal data protection and usability.

To overcome these barriers, implementers must address not only the technical and governance challenges but also prioritize clear and compelling narratives that highlight the benefits of decentralized identities to stakeholders, governments, and individuals alike. Only then can these systems realize their full potential.

By categorizing these barriers and analyzing their implications, this section and WEF's paper, *Reimagining Digital ID* [28], highlight the multifaceted nature of the challenges facing the integration of decentralized identities into CBDC systems. Addressing these issues will require coordinated efforts from policymakers, technologists, and other stakeholders to create resilient, inclusive, and user-friendly systems.

3.3 Current Developments

As CBDCs gain prominence globally, this section examines the current state of adoption and ongoing projects focused on integrating decentralized identities into CBDC frameworks. The exploration of these developments reflects the growing interest among central banks to modernize payment systems while addressing challenges such as privacy, security, and inclusivity.

Central banks have significantly accelerated their CBDC efforts in recent years, as illustrated in Figure 3.3, which highlights the cumulative progress of retail and wholesale CBDC projects. The chart shows an increase in CBDC projects since 2017, driven by multiple factors such as the declining use of cash, the rise of private digital currencies, and the need for more efficient cross-border payment systems [36]. In the latter part of 2020, central banks representing approximately one-fifth of the global population indicated that they were intending to issue CBDCs in the near future [38]. Retail CBDCs, in particular, are advancing at a faster pace than wholesale CBDCs, reflecting their potential to enhance financial inclusion and provide a digital alternative to cash for everyday transactions.

Figure 3.3 provides a visual representation of the growing commitment to CBDC development worldwide. It categorizes projects by their stage, research, pilot, or live, and distinguishes between retail and wholesale applications. The steady rise in retail CBDC initiatives reflects a strong demand for secure and accessible digital payment options among the general public, while wholesale CBDC projects remain crucial for enhancing efficiency in interbank settlements.

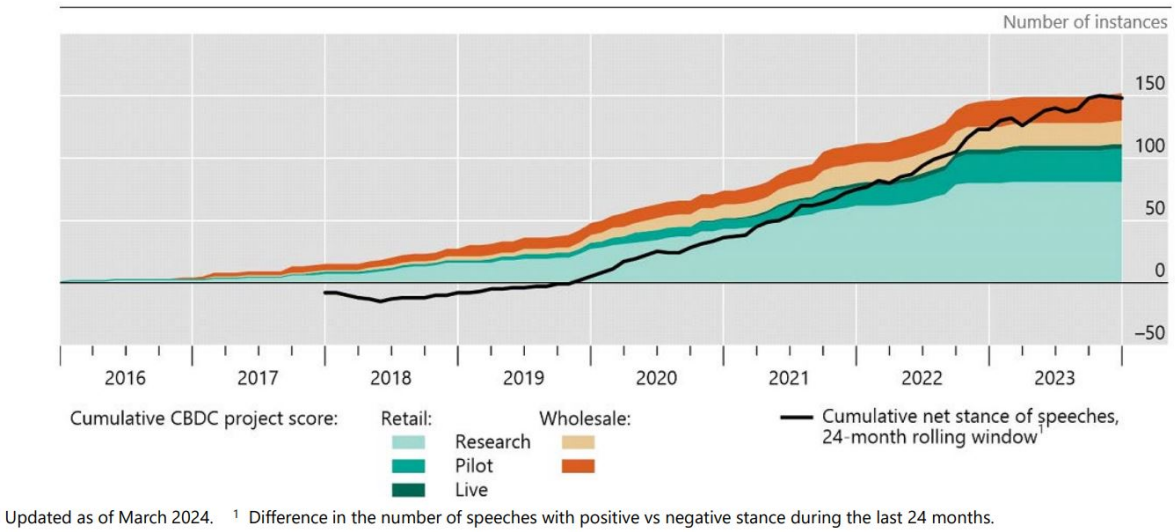


Figure 3.3: Number of central banks working on CBDC projects (Source: *Auer, Cornelli, and Frost (2023)* [37])

The first subsection, 3.3.1, goes into specific CBDC projects and outlines the technical frameworks behind them. This discussion includes key examples of successfully implemented retail CBDCs and highlights ongoing pilot projects. Furthermore, it explores the concerns and barriers that these projects face, including technical scalability and public acceptance.
The second subsection, 3.3.2, focuses on the integration of decentralized identity frameworks into CBDCs. Decentralized identities, such as those based on SSI principles, are increasingly recognized as essential for ensuring privacy and user control in digital transactions. This subsection explores projects and academic research that aim to leverage decentralized identities to strengthen CBDC systems, emphasizing the technical and governance innovations required for their successful adoption.
By examining these developments, this section aims to provide an overview of the current state of CBDC adoption and the innovations shaping their future. The analysis offers insights into how central banks are addressing the technical, policy, and governance challenges associated with CBDCs.

3.3.1 Adoption of CBDC

The global adoption of CBDCs has accelerated in recent years, with four retail CBDCs successfully implemented. These include the Sand Dollar in the Bahamas, DCash in the Eastern Caribbean Currency Union (ECCU), the eNaira in Nigeria, and Jam-Dex in Jamaica. These projects provide valuable insights into the technical frameworks and challenges associated with retail CBDC adoption.
The Sand Dollar (Bahamas): The Sand Dollar, officially launched in October 2020, is the world's first operational retail CBDC [39]. It is the digital version of the Bahamian dollar and is issued by the Central Bank of The Bahamas (CBB) through authorized financial institutions (AFIs). This initiative emerged as part of the Bahamian Payments System Modernisation Initiative (PSMI), targeting greater financial inclusion, particularly for residents on remote islands where traditional banking services are scarce [1] .
The Sand Dollar employs a token-based system supported by DLT and maintains a tiered wallet structure. Tier I wallets require no official identification to open a digital account, catering to low-value transactions. Tier II or premium wallets are also for individuals, but do require some customer due diligence, which can be simplified but risk-based. Tier II

wallets require presentation of government issued ID. Tier III wallets, designed for businesses, impose no preset transaction limits but need comprehensive KYC compliance [40]. Offline functionality is a notable feature, enabling transactions even when communications are disrupted, with wallet balances synchronizing once connectivity is restored.

Despite its potential, adoption of the Sand Dollar has been constrained by limited public awareness and skepticism about digital payment systems. The Sand Dollar currently accounts for less than 1% of the total currency in circulation in the Bahamas. One of the key challenges is that the Sand Dollar does not yet offer any clear and compelling advantages over existing payment methods. Additionally, public concerns persist that the Sand Dollar could potentially lead to increased government surveillance [41].

DCash (Eastern Caribbean Currency Union): The Eastern Caribbean Central Bank (ECCB) launched the DCash pilot in March 2021, becoming the first monetary union to implement a retail CBDC. DCash is issued and managed by the ECCB and is designed to reduce cash dependency, lower transaction costs, and increase financial inclusion across the ECCU's eight member states [42].

DCash operates as a token-based CBDC built on DLT technology provided by Bitt Inc [43]. Unlike the Sand Dollar, DCash does not currently support offline transactions, relying instead on internet connectivity for real-time settlement. The system includes both value-based wallets, which allow unbanked users to participate in the digital economy, and registered wallets, which are linked to bank accounts. Identity verification follows a risk-based approach, with KYC requirements tailored to wallet types [1].

The ECCB addresses privacy concerns by ensuring that all processes comply with the General Data Protection Regulation (GDPR), other international data protection standards, and relevant local and regional laws [1]. According to the ECCB, personal data is exclusively accessible to the user's financial institution and is transmitted through encrypted channels to safeguard confidentiality. Additionally, any personal data stored on disk is encrypted and securely housed in a protected facility, further reinforcing the privacy and security of the system [42].

It is worth noting, however, that despite these assurances, data is still stored and managed by an intermediary. This reliance on intermediaries may raise concerns among individuals who prioritize complete control over their personal data and prefer decentralized data storage solutions.

eNaira (Nigeria): On October 25, 2021, the Central Bank of Nigeria (CBN) launched the eNaira, Nigeria's retail CBDC [44]. The eNaira is intended to address several key policy objectives, including enhancing financial inclusion, improving access to central bank money, and increasing the efficiency and resilience of payment systems. Furthermore, the eNaira is intended to reduce the costs associated with cross-border payments and facilitate more affordable remittances to Nigeria [45]

The eNaira employs a two-tier distribution model: the CBN oversees issuance and minting through its Digital Currency Management System (DCMS), while financial institutions manage currency holdings and distribution using Treasury Wallets linked to the DCMS. The eNaira platform includes several wallet types tailored to different user groups. At the highest level, the Stock Wallet, maintained by the CBN, acts as the reservoir for all minted eNaira. Financial institutions utilize Treasury Wallets, which are subdivided into Branch Wallets for local operations. End users can access either Basic Speed Wallets for retail payments or Merchant Speed Wallets designed for business transactions [45].

From a technological standpoint, the eNaira relies on permissioned DLT, where network nodes are controlled by intermediaries such as financial institutions. The eNaira employs a tiered KYC framework to regulate transaction and balance limits. Lower-tier wallets, which require minimal identification, are accessible to unbanked individuals and have smaller transaction limits, while higher-tier wallets necessitate extensive KYC verification

and allow greater usage capacity. For a complete overview of the various tiered wallet systems of eNaira, see figure 3.4.

eNaira: Tiered Wallet System				
Tier	Client category	Requirement to open eNaira wallet	Identity test	Illustrative Limits
0	Retail (including people without bank account)	Phone number	No identify information required except for phone number	. Daily transaction limit (N20,000) . Balance limit (N120,000)
1	Retail (including people without bank account)	Phone number (national id number verified)	Basic identity information (e.g., photo, name, date of birth); no evidence required; no verification required	. Daily transaction limit (N50,000) . Balance limit (N300,000)
2	Retail (people with bank account)	Bank verification number (BVN)	Basic identity information (e.g., photo, name, date of birth); evidence required for submitted information; customer to be verified through official databases	Daily transaction limit (N200,000) Balance limit (N500,000)
3	Retail (people with bank account)	Bank verification number (BVN)	Full identify information and evidence (including proof of address and physical presence in the address) in pursuant to CBN's AML/CFT Regulation 2009. Risk-based verification done.	Daily transaction limit (N1,000,000) Balance limit (N5,000,000)
Merchant		Existing bank account, TIN, BVN of MD/CEO, email address, business certificate	Full KYC requirement in pursuant to CBN's AML/CFT Regulation 2009.	No limit

Figure 3.4: eNaira: Tiered Wallet System (Source: *Ree (2023)* [44])

The CBN prioritizes universal access to the eNaira by issuing digital identification for unbanked individuals to facilitate their inclusion in the digital financial ecosystem. Wallet caps and transaction limits are designed to encourage the use of eNaira for small-scale retail payments, avoiding competition with traditional bank deposits. Furthermore, the eNaira operates with a 0% interest rate, reinforcing its role as a payment mechanism rather than a savings tool [45].

In regard to anonymity, the eNaira system is designed with complete traceability in mind. Even the lower-tiered wallets need the input of a bank verification number or, at the very least, a verified phone number for the account setup process. This decision reflects a commitment to security and regulatory compliance, although it does result in a reduction in the privacy of users. Users are apprehensive about the CBN's ability to monitor all transactions, leading to fears of potential misuse of personal data and financial surveillance [46].

Jam-Dex (Jamaica): In 2022, Jamaica introduced Jam-Dex, its retail CBDC, to modernize the country's payment systems and promote financial inclusion. Jam-Dex operates on a centralized architecture, ensuring oversight and regulatory compliance by the Bank of Jamaica (BOJ). The system is designed to enable fast and secure transactions while offering offline payment capabilities for regions with limited connectivity [47].

Jam-Dex stands out for its emphasis on reducing barriers to entry, enabling seamless access to the digital currency for individuals and businesses alike. Users can register and activate wallets through authorized financial institutions with minimal identification requirements, aligning with KYC standards while promoting inclusivity . The system is designed to facilitate fast, secure, and low-cost transactions, fostering greater economic participation among unbanked and underbanked populations [48].

Jam-Dex faces significant challenges, including low adoption rates, limited public trust, and barriers to integration within the broader financial ecosystem. A key issue is the lack of public awareness and understanding of its purpose and benefits, leading to hesitance and mistrust of government-issued digital currencies. Many Jamaicans prefer the perceived stability of cash or private digital platforms. Accessibility and infrastructure pose additional barriers. Low levels of digital literacy further complicate adoption, as many users lack the confidence or skills to use digital wallets effectively [49].
These four operational CBDCs highlight the potential of retail CBDC adoption but also underscore significant challenges, including technical scalability, public trust, and effective user education. A critical limitation across all four projects is their reliance on centralized models for managing user identification. Each project depends on intermediaries and central banks to handle the private data required for identification, raising privacy concerns.
These concerns could be mitigated through the integration of decentralized identity frameworks, which empower users to manage their own digital identities independently. By reducing reliance on centralized data storage and enabling greater user control over personal information, decentralized identities could enhance privacy, build trust, and address a key barrier to broader CBDC adoption. Addressing these systemic issues through such innovations will be essential for shaping the future success and public acceptance of CBDCs.

3.3.2 Projects Using Decentralized Identities

The European Union (EU) has recognized the importance of building trust in online interactions as a cornerstone for societal and economic development [50]. In pursuit of this objective, the EU has introduced the European Digital Identity Regulation, commonly referred to as eIDAS 2.0 (Regulation (EU) 2024/1183) [51]. This regulation, which entered into force in May 2024, aims to establish a comprehensive framework for digital identities across member states [52].
A central component of eIDAS 2.0 is the European Digital Identity Wallet. This digital wallet is designed to enable EU citizens, residents, and businesses to securely identify themselves and share personal information for accessing both public and private services online and offline throughout the EU. The wallet allows users to prove their identity, share digital documents, and confirm specific personal attributes, such as age, without disclosing unnecessary personal details. Importantly, users maintain full control over the data they share and can determine who has access to their information [51].
The regulation mandates that EU member states provide these digital identity wallets to their citizens within 24 months following the adoption of implementing acts that outline technical specifications and certification processes. .
The overarching goal of eIDAS 2.0 is to enhance digital trust and facilitate seamless access to essential public services. The EU has set an ambitious target: by 2030, 80% of EU citizens should be able to access key public services securely using a digital identity. This initiative is part of the EU's broader digital agenda, which seeks to create a secure digital space, ensure fair competition in digital markets, and strengthen Europe's digital sovereignty [50].
The eIDAS 2.0 framework addresses key shortcomings of the original *regulation 910/2014* [26] by introducing a more flexible and user-centric approach to digital identity. Unlike the earlier system, which required a rigid, singular ID that disclosed extensive personal information indefinitely, eIDAS 2.0 incorporates a self-sovereign identity model. This innovative structure allows individuals to maintain full control over their identifying information, enabling them to selectively share only the necessary details for specific trans-

actions. This paradigm shift empowers users while supporting both public and private digital ecosystems [53].

By using cryptographic proofs, SSIs enable verification of specific elements of a person's identity without revealing unrelated personal data. This approach not only meets the high standards of authenticity demanded by the eIDAS framework but also ensures consumer privacy [53]. Integrating these principles with decentralized technologies, such as blockchain, elevates eIDAS 2.0 to a leading example of privacy and security in digital identity systems.

However, as pointed out by *Johnson (2022)* [53] challenges remain. Third parties could still collect and store whatever data they can access under this system, potentially undermining the privacy and autonomy SSIs aim to protect. To address this, safeguards must be established to ensure that user data remains encrypted and inaccessible in its raw form. Information should function solely as a key for verification, not as human-readable content. Zero-Knowledge proofs play a critical role here by enabling verification without exposing the underlying information. These proofs provide absolute assurance of identity legitimacy without giving entities the opportunity to extract or exploit user data.

Biometric identifiers, such as fingerprints and iris scans, are another cornerstone of eIDAS 2.0's security framework. When paired with robust privacy protections, these unique physical markers ensure that only authorized individuals can access their SSI [53]. This combination of selective disclosure, cryptographic safeguards, and biometrics establishes a new standard for digital identity systems, offering enhanced levels of privacy, security, and user control.

Alongside these advancements in the eIDAS 2.0 framework, the European Commission unveiled a legislative package to introduce their own CBDC, the digital euro, in June 2023 [54]. This initiative aims to complement physical currency while setting out the rules and conditions governing its use, aligning with broader efforts to modernize the EU's digital financial infrastructure [50].

The combination of the digital euro initiative with the eIDAS 2.0 framework offers a promising vision for the creation of a secure and inclusive digital ecosystem within the European Union. The integration of a CBDC with a robust SSI framework has the potential to enhance the security, efficiency, and user-centricity of digital financial transactions. By leveraging the privacy-preserving principles of eIDAS 2.0, the digital euro could ensure that users maintain control over their identity and personal data, thereby addressing one of the key concerns associated with digital currencies.

This concept of connecting the eIDAS standards with a CBDC framework has already been proposed by *Adams et al. (2021)*, as was shown in section 3.2.2. Their findings demonstrated that such integration could enhance the efficiency of verification processes, reduce the incidence of fraud, and improve the protection of user privacy. Given the establishment of the eIDAS 2.0 framework and the legislative basis for the digital euro, the EU is well-positioned to lead the way in this regard.

In the future, the combination of these two initiatives may result in the creation of a unified digital ecosystem that promotes trust, safeguards user privacy, and guarantees accessibility for all EU citizens. Such an approach would not only reinforce Europe's digital autonomy but also establish a global standard for the harmonious coexistence of digital identities and CBDCs in a secure and user-centric manner.

Several jurisdictions beyond the European Union are also actively researching and implementing decentralized identity frameworks.

In December 2023, the Chinese Ministry of Public Security, in collaboration with the Blockchain-based Service Network (BSN), China's national blockchain scheme, officially launched the China Real-Name Decentralized Identifier System (China RealDID), a national-level decentralized identifier system [55; 56]. China RealDID is a blockchain-based real-name identity verification system. It offers real-name verification, data encryption, secure

logins, business identity checks, and personal ID services. It allows Chinese residents to access online services using DID addresses and private keys [55]. Simultaneously, China continues to advance research and pilot programs for the e-Yuan, its CBDC, which has already been rolled out in 29 cities [57]. The integration of decentralized identity frameworks like China RealDID with the e-Yuan could be explored to enhance security and align the digital currency with the country's broader digital transformation goals.
Another jurisdiction that is engaged in the development of decentralized identity frameworks is South Korea. In this context, a public/private consortium has been established with the specific objective of advancing the field of decentralized identity [58].
The integration of decentralized identity frameworks with financial systems shows the potential of such technologies. Decentralized identities can redefine user privacy, security, and control, particularly when coupled with CBDCs. The efforts discussed in this section illustrate a recognition of the necessity for secure digital ecosystems, setting the stage for a bigger global adoption. The integration of decentralized identities with CBDCs represents a shift in how trust and autonomy could be managed in the digital era.

3.4 Targeting the Barriers

In section 3.2.3, we identified key obstacles to the integration of decentralized identity frameworks with CBDCs, encompassing technical, policy-related, governance, and implementation challenges. Addressing these barriers is crucial for ensuring the successful adoption and functionality of both CBDCs and decentralized identities.
This section explores strategies to overcome these challenges. First, I will focus on creating the right environment for the implementation of CBDCs and decentralized identity systems. Furthermore, section 3.4.2 will present specific policy recommendations and technical solutions aimed at addressing the identified barriers. These suggestions will provide a roadmap for stakeholders, including policymakers, financial institutions, and technology providers, to enable a seamless and secure integration of decentralized identity frameworks into digital financial ecosystems.
By tackling these barriers systematically, the potential of decentralized identities to enhance privacy, security, and user control within CBDC systems can be realized, paving the way for a more inclusive digital economy.

3.4.1 Creating the Right Environment

Establishing the right environment for implementing CBDCs and decentralized identity frameworks requires a comprehensive, multi-faceted approach. The REDI Framework, Regulation, Education, Design, and Incentives, presented by the IMF Fintech Note *Central Bank Digital Currency Adoption: Inclusive Strategies for Intermediaries and Users* [29] provides central banks and policymakers with a structured methodology to address the key challenges and opportunities involved in CBDC adoption.
Regulation: A robust legal framework is essential to promote adoption and trust in CBDCs. Clear guidelines that define intermediary participation, ensure user data privacy, and establish minimum quality standards for services are pivotal. For instance, regulations should mandate high-security protocols to protect user data while accommodating a tiered approach to KYC to balance accessibility and compliance. In addition, granting CBDCs legal tender status can encourage both consumer and merchant acceptance.
Education: Public awareness campaigns and user education initiatives play a crucial role in dispelling misconceptions and building trust. Central banks and policymakers should design educational content that emphasizes the benefits of CBDCs and decentralized identities, such as enhanced security, privacy, and efficiency. Tailoring messages to

specific stakeholder groups, consumers, merchants, and intermediaries, ensures effective communication and builds confidence in these technologies. For example, the deployment of localized outreach programs, such as "ambassador programs", can be instrumental in engaging communities directly.

Design: User-centric design is critical to the success of CBDCs and decentralized identity systems. Simplified interfaces, multilingual support, and compatibility with existing payment systems can make these technologies accessible to a diverse user base. Features like offline functionality and interoperability with private financial tools enhance usability and inclusivity, particularly for underserved populations. Central banks should ensure that these systems integrate seamlessly into existing financial ecosystems, minimizing the need for costly infrastructure upgrades by merchants.

Incentives: Financial and non-financial incentives are necessary to encourage stakeholder participation. For instance, intermediaries such as financial institutions and merchants may need financial support to cover integration costs or revenue-sharing mechanisms to offset potential revenue losses. Simultaneously, consumers could be incentivized through benefits such as sign-up bonuses in which new users receive a one-time deposit of CBDC. In Jamaica, the first 100,000 citizens who registered for a Jam-Dex wallet after April 1, 2022, were incentivized with a deposit of JD$2,500 (approximately CHF 14 or US$16) [59]. As a non-financial incentive, central banks can offer white-label CBDC wallet solutions that intermediaries can customize and brand as their own. This approach combines the technological reliability of a central bank-backed system with the strategic branding flexibility needed to attract and retain customers. By using technology provided by the central bank, intermediaries can avoid significant development costs while benefiting from the added credibility and trust that come with central bank affiliation.

By addressing these key areas, the REDI Framework by *IMF (2024* [29] provides a roadmap for creating the right environment for CBDCs and decentralized identities. Successful adoption hinges not only on the technical readiness of these systems but also on the collective efforts of stakeholders to foster trust, accessibility, and functionality. Central banks and policymakers must work collaboratively to ensure these innovations meet the diverse needs of their populations.

3.4.2 Policy suggestions and Technical Solutions

This section builds on the barriers discussed in Section 3.2.3 and follows the recommendations provided in the WEF report, *Reimagining Digital ID* [29], to propose strategies for the successful implementation of decentralized identities. The recommendations are categorized into technical, policy, governance, and implementation measures.

Technical Recommendations: To address technical barriers, stakeholders must prioritize investments in developing the technologies underlying decentralized identity systems. This includes closing funding gaps to support innovations such as identity recovery mechanisms and secure revocation protocols. Stakeholders can also benefit from promoting collaboration across ecosystems to reduce costs and risks.

Another critical element is the alignment of technical standards. Engaging with public-private partnerships, such as the OpenWallet Foundation and W3C, can accelerate the creation of interoperable digital wallets and verifiable credential standards. Sharing lessons learned from pilot projects and adopting a multi-ecosystem approach, enabling different networks of verifiers and issuers to operate effectively, can also help scale decentralized identity systems.

To overcome challenges in change and process management, stakeholders should prioritize talent development. This includes creating training and certification programs focused on decentralized ID technologies, which serve both as a mechanism for workforce development and as an incentive for individuals to engage in the field. Cross-organizational

collaboration in open-source, open-standards, and co-development organizations can further bolster technical skills and foster a culture of innovation. Another important area is design. Collaborating with experienced human-centric design researchers and usability experts can significantly enhance user-interface and user-experience designs. Simplifying user-management processes for ID credentials is a prime example of how design thinking can improve accessibility and adoption. Intuitive designs that prioritize simplicity can empower users to manage their digital identities with ease, reducing friction in everyday interactions.

Policy Recommendations: Policymakers must evaluate and adapt existing regulatory frameworks to align with the objectives of decentralized identities. This includes removing systemic barriers, such as laws that restrict the use of reusable credentials, and exploring enabling regulations to define requirements for digital wallets and validators. For example, auditing and certification processes can ensure that trusted validators meet established standards.

Governments should incentivize the development of privacy-enhancing technologies through policies that promote data portability and interoperability. Initiatives like GDPR have demonstrated how regulations can encourage innovation while protecting user privacy. Transitional mechanisms, such as regulatory sandboxes, allow governments to test new technologies while promoting innovation and collaboration between stakeholders.

Government stakeholders can play an important role in promoting collaboration between the public and private sectors by using existing initiatives to establish effective communication channels. These efforts are crucial for ensuring a robust exchange of information between government bodies and industry players. By clearly articulating the benefits and potential risks associated with decentralized identity systems, governments can build understanding and support among lawmakers and their constituencies.

Additionally, international fora offer valuable platforms for sharing experiences, best practices, and lessons learned across jurisdictions, enabling a collective approach to addressing challenges and refining implementation strategies.

Governance and Implementation Recommendations: Effective governance and implementation frameworks are essential for promoting trust and accountability in decentralized identity systems. Clear communication of benefits and risks is a critical governance measure. Public awareness campaigns should focus on explaining how decentralized identity systems enhance privacy, user control, and efficiency, while countering misinformation and addressing concerns about misuse.

Practical implementation strategies must target user needs and prioritize inclusivity. Developing decentralized identities with clear use cases, such as education credentials, can drive adoption. Stakeholders should also design systems that are accessible to users with minimal digital literacy, bridging the digital divide by offering affordable and easy-to-use tools.

Mitigating exclusion and marginalization requires localized research to understand community-specific barriers. Governments and stakeholders should also establish ethical standards to prevent coerced consent and ensure fairness in system deployment. Additionally, leveraging trusted wallets and hybrid systems that integrate analog and digital approaches can ensure wider accessibility.

By combining technical advancements, adaptive policies, robust governance, and inclusive implementation strategies, stakeholders can create a sustainable ecosystem for decentralized identities. These measures are not only essential for overcoming existing barriers but also for realizing the transformative potential of decentralized identity frameworks in digital financial systems, especially in CBDC.

3.5 Conclusions and Future Outlook

The evolution of CBDCs represents a transformative shift in the global financial landscape, with the potential to enhance financial inclusion, streamline payment systems, and modernize monetary policies. However, as this report has explored, the successful implementation of CBDCs faces numerous challenges, particularly around privacy and security. These issues are deeply rooted in the reliance on centralized systems for identity management and the handling of sensitive user data, which raises concerns about the potential for data misuse and surveillance.

Decentralized identity frameworks offer a promising solution to these challenges. By enabling users to control their own digital identities, these frameworks can enhance privacy, create trust, and align with the goals of financial inclusion. Theoretical models, such as those proposed by *Adams et al. (2021)* [9], alongside initiatives like the European Union's eIDAS 2.0, demonstrate the feasibility of integrating decentralized identities into CBDC ecosystems. While these developments mark significant progress, practical implementations remain limited, leaving much work to be done to translate theory into practice.

The creation of an appropriate environment for the adoption of CBDCs and decentralized identity requires the coordinated efforts of policymakers, financial institutions, and technology providers. As detailed in this report, a range of targeted policy recommendations and technical solutions, including public education campaigns and incentives for end-users and merchants, can address the barriers to adoption. Furthermore, it must be emphasized that governments must prioritize transparency and collaboration in order to build public trust. These systems must be inclusive, secure, and user-centric.

Looking to the future, the growing interest in decentralized identity frameworks across jurisdictions offers a hopeful outlook. Initiatives in the EU, China, and South Korea illustrate the global recognition of the necessity for secure and privacy, preserving digital ecosystems. The combination of decentralized identities with CBDCs offers a promising path forward for central banks and policymakers seeking to enhance the efficiency of financial systems while simultaneously empowering individuals by giving them greater control over their personal data.

In conclusion, the integration of decentralized identity frameworks into CBDC systems represents an important opportunity to address the key challenges of privacy and security in digital financial ecosystems. As governments and institutions continue to innovate, the successful implementation of these frameworks could establish a global standard for the harmonious coexistence of financial inclusion, user autonomy, and technological advancement. By addressing the obstacles outlined and embracing the potential of decentralized identities, a future of secure, efficient, and inclusive digital economies is within reach.

Bibliography

[1] D. W. Patrick and T. Lyle, "Central bank digital currency: Caribbean pathways," *ECB Legal Conference 2021, Continuity and change - How the Challenges of Today Prepare the Ground for Tomorrow*, p. 340, 2022.

[2] J. Parkin, "Cashless payment is booming thanks to coronavirus - so is financial surveillance," *The Conversation*, 2020. https://theconversation.com/cashless-payment-is-booming-thanks-to-coronavirus-so-is-financial-surveillance-145179.

[3] Allianz Global Investors, "COVID-19 and the Cashless Society," 2020. https://ch.allianzgi.com/de-de/en-insights/thematic-investing/covid-19-and-the-cashless-society.

[4] E. Feyen, J. Frost, L. Gambacorta, H. Natarajan, and M. Saal, "Fintech and the digital transformation of financial services: implications for market structure and public policy," *BIS reports*, 2021, Bank for International Settlements.

[5] N. Maynard, "Digital Wallets Adoption Driven by COVID-19 Pandemic," *Juniper Research*, 2020. https://www.juniperresearch.com/resources/blog/digital-wallets-adoption-driven-by-covid-19-pandem.

[6] Bank for International Settlements, "Central bank digital currencies: foundational principles and core features," Bank for International Settlements, 2020. https://www.bis.org/publ/othp33.htm.

[7] M. Salmony, "Do we really need another dollar, euro, pound or yuan? How to create the right ecosystem for a successful central bank digital currency," *Journal of Payments Strategy & Systems*, vol. 17, March 2023. doi: 10.69554/PRSE6720.

[8] R. Auer, G. Cornelli, and J. Frost, "Rise of the Central Bank Digital Currencies: Drivers, Approaches, and Technologies," *BIS Working reports*, no. 880, Bank for International Settlements, August 2020, updated March 2024. https://www.bis.org/publ/work880.htm.

[9] M. Adams, L. Boldrin, R. Ohlhausen, and E. Wagner, "An integrated approach for electronic identification and central bank digital currencies," *Journal of Payments Strategy & Systems*, vol. 15, no. 3, pp. 287–304, 2021.

[10] Reserve Bank of India, "Report on Central Bank Digital Currency," 2021. https://rbi.org.in/Scripts/PublicationReportDetails.aspx?UrlPage=&ID=1218.

[11] A. Abi Karam, "Central Bank Digital Currency (CBDC) & Blockchain: The Future of Payments," IBM, August 17, 2023. https://www.ibm.com/think/topics/blockchain-for-cbdc.

[12] Bank for International Settlements, "Central Bank Digital Currencies: Foundational Principles and Core Features," BIS report, 2020. ISBN: 978-92-9259-427-5. https://www.bis.org/publ/qtrpdf/r_qt2003j.htm.

[13] L. V. Schumacher, *Decoding Digital Assets: Distinguishing the Dream from the Dystopia in Stablecoins, Tokenized Deposits, and Central Bank Digital Currencies*, Springer Nature, 2024.

[14] Bank for International Settlements, "BIS Quarterly Review: International Banking and Financial Market Developments," 2020. https://www.bis.org/publ/qtrpdf/r_qt2003.htm.

[15] W. Bossu, M. Itatani, C. Margulis, A. Rossi, H. Weenink, and A. Yoshinaga, "Legal aspects of central bank digital currency: Central bank and monetary law considerations," *IMF Working Paper*, 2020.

[16] Bank for International Settlements, "BIS Annual Economic Report 2021," June 2021. https://www.bis.org/publ/arpdf/ar2021e.htm.

[17] G. Soderberg, J. Kiff, M. Bechara, S. Forte, K. Kao, A. Lannquist, T. Sun, H. Tourpe, and A. Yoshinaga, "How Should Central Banks Explore Central Bank Digital Currency? A Dynamic Decision-Making Framework," *IMF Fintech Note*, no. 2023/008, International Monetary Fund, Washington, DC, 2023.

[18] E. A. Opare and K. Kim, "A Compendium of Practices for Central Bank Digital Currencies for Multinational Financial Infrastructures," *IEEE Access*, vol. 8, pp. 110810-110835, 2020. doi: 10.1109/ACCESS.2020.3001970.

[19] M. Bijlsma, C. van der Cruijsen, N. Jonker, and J. Reijerink, "What Triggers Consumer Adoption of CBDC?" *De Nederlandsche Bank Working report*, no. 709, April 2021.

[20] T. Mancini-Griffoli, M. S. Martinez Peria, I. Agur, A. Ari, J. Kiff, A. Popescu, and C. Rochon, "Casting Light on Central Bank Digital Currency," *IMF Staff Discussion Note*, SDN/18/08, International Monetary Fund, November 2018.

[21] R. Auer and R Boehme, "The technology of retail central bank digital currency", *BIS Quarterly Review*, March 2020.

[22] Bank for International Settlements, "Central Bank Digital Currencies: System Design and Interoperability," BIS report, 2021. ISBN: 978-92-9259-510-4. https://www.bis.org/publ/othp42_system_design.pdf.

[23] European Central Bank, "Privacy and the Digital Euro," European Central Bank, 2024. https://www.ecb.europa.eu/euro/digital_euro/features/privacy/html/index.en.html.

[24] D. Birch, "National Digital ID is a Foundation for CBDC," Forbes, April 26, 2023. https://www.forbes.com/sites/davidbirch/2023/04/26/national-digital-id-is-a-foundation-for-cbdc/.

[25] Johnson, Alastair, "Self-Sovereign Decentralized Identity Is Key to Retail CBDC Adoption," Forbes, August 7, 2024. https://www.forbes.com/sites/alastairjohnson/2024/08/07/self-sovereign-decentralized-identity-is-key-to-retail-cbdc-adoption/.

[26] European Union, "Regulation (EU) No 910/2014: Electronic Identification and Trust Services for Electronic Transactions in the Internal Market,", September 23, 2014. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2014.257.01.0073.01.ENG.

[27] Regulatory Technical Standards on Strong Customer Authentication & Common and Secure Communication, "Commission Delegated Regulation (EU) 2018/389 of 27 November 2017", March 2018. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32018R0389.

[28] World Economic Forum, "Reimagining Digital ID", June 7, 2023. https://www.weforum.org/publications/reimagining-digital-id/.

[29] International Monetary Fund, "Central Bank Digital Currency Adoption: Inclusive Strategies for Intermediaries and Users," IMF Fintech Notes, September 20, 2024. https://www.imf.org/en/Publications/fintech-notes/Issues/2024/09/21/Central-Bank-Digital-Currency-Adoption-Inclusive-Strategies-for-Intermediaries-and-Users-555118.

[30] Poehn, D., Grabatin, M., and Hommel, W., "eID and Self-Sovereign Identity Usage: An Overview," Electronics, vol. 10, no. 22, p. 2811, 2021. https://doi.org/10.3390/electronics10222811.

[31] Decentralized ID.com, "Verifiable Credentials - Literature, Comparisons, Explainer (W3C)," https://decentralized-id.com/web-standards/w3c/verifiable-credentials/.

[32] Becker, K., Serota, L., Scuri, S., and Wang, T., "UX in Cryptocurrency: An Overview of User Experience in Cryptocurrency Applications," CRADL Report, August 2022. https://docs.google.com/presentation/d/1s2OPSH5sMJzxRYaJSSRTe8W2iIoZx0PseIV-WeZWD1s/edit#slide=id.p.

[33] Alrehaili, A., Namoun, A., and Tufail, A., "A Comparative Analysis of Scalability Issues within Blockchain-Based Solutions in the Internet of Things," International Journal of Advanced Computer Science and Applications, vol. 12, no. 9, 2021. https://thesai.org/Downloads/Volume12No9/Paper_55-A_Comparative_Analysis_of_Scalability_Issues.pdf.

[34] Sandman, J., "It's Time for IDs to Go Digital," New America, August 24, 2021. https://www.newamerica.org/the-thread/its-time-for-ids-to-go-digital/.

[35] Brennan Center for Justice, "Citizens Without Proof: A Survey of Americans' Possession of Documentary Proof of Citizenship and Photo Identification," November 2006. https://www.brennancenter.org/sites/default/files/legacy/d/download_file_39242.pdf.

[36] Auer, R., Cornelli, G., and Frost, J., "Rise of the Central Bank Digital Currencies," International Journal of Central Banking, vol. 19, no. 4, pp. 185-214, 2023. https://www.ijcb.org/journal/ijcb23q4a5.pdf.

[37] Auer, R., Cornelli, G., and Frost, J., "Updates to BIS Working Paper No. 880: Rise of the Central Bank Digital Currencies," March 2024. https://www.bis.org/publ/work880_updates_mar2024.pdf.

[38] Boar, C., and Wehrli, A., "Ready, steady, go? - Results of the third BIS survey on central bank digital currency," BIS Papers No. 114, Bank for International Settlements, January 2021. https://www.bis.org/publ/bppdf/bispap114.pdf.

[39] Dorst, S., "Digital Dollars for Online Tea," IMF Finance & Development, March 2021. https://www.imf.org/external/pubs/ft/fandd/2021/03/fighting-pandemic-disruption-with-innovation-dorst.htm.

[40] Central Bank of The Bahamas, "Consumer-Centric Aspects of the Proposed Regulations for the Bahamian Digital Currency," March 2021. https://www.centralbankbahamas.com/viewPDF/documents/2021-03-26-12-00-35-PSD-Policy-Paper-on-Consumers-Issues.pdf.

[41] Reuters, "Bahamas to regulate banks to offer cbank digital currency," July 1, 2024. https://www.reuters.com/technology/bahamas-regulate-banks-offer-cbank-digital-currency-2024-07-01/.

[42] Eastern Caribbean Central Bank, "Frequently Asked Questions,", 2024. https://www.eccb-centralbank.org/frequently-asked-questions.

[43] Bitt Inc., "Building the world's first CBDC in a currency union,", 2024. https://www.bitt.com/success-stories/dcash.

[44] Ree, J., "Nigeria's eNaira, One Year After," IMF Working Paper WP/23/104, Washington, DC: International Monetary Fund, 2023. https://www.imf.org/en/Publications/WP/Issues/2023/05/16/Nigerias-eNaira-One-Year-After-533487

[45] Bank for International Settlements, "Central Bank Digital Currencies in Africa," BIS Paper, 2023. https://www.bis.org/publ/bppdf/bispap128.pdf.

[46] The Conversation, "eNaira: Nigeria's Digital Currency Has Had a Slow Start - What's Holding It Back?" July 19, 2023. https://theconversation.com/enaira-nigerias-digital-currency-has-had-a-slow-start-whats-holding-it-back-209470.

[47] Bank of Jamaica, "Central Bank Digital Currency (CBDC) FAQs Booklet," August 2022. https://boj.org.jm/wp-content/uploads/2022/08/CBDC-FAQs-Booklet.pdf.

[48] eCurrency, "eCurrency Awarded Best Financial Inclusion Initiative in Digital Currency," accessed November 20, 2024. https://www.ecurrency.net/post/ecurrency-awarded-best-financial-inclusion-initiative-in-digital-currency.

[49] ICT Pulse, "Roadblocks to a Successful Digital Marketplace for Jam-Dex: 5 Critical Issues to Tackle," May 12, 2023. https://ict-pulse.com/2023/05/roadblocks-to-a-successful-digital-marketplace-for-jam-dex-5-critical-issues-to-tackle/.

[50] European Parliament, "Digital Agenda for Europe," April 2024. https://www.europarl.europa.eu/factsheets/en/sheet/64/digital-agenda-for-europe.

[51] Cyber Risk GmbH, "The European Digital Identity Regulation," 2024. https://www.european-digital-identity-regulation.com/.

[52] European Union, "Regulation (EU) 2024/1183: European Digital Identity," April 2024. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1183.

[53] Johnson, A., "eIDAS 2.0 Turns to Self-Sovereign Identification to Bring Users Ownership and Control," Forbes, July 5, 2022. https://www.forbes.com/sites/alastairjohnson/2022/07/05/eidas-20-turns-to-self-sovereign-identification-to-bring-users-ownership-and-control/.

[54] European Commission, "Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the establishment of the digital euro," COM/2023/369 final, June 28, 2023. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52023PC0369.

[55] Business Insider, "China's Ministry of Public Security Launches Blockchain-Based Real-Name Decentralized Identifier System," December 12, 2024. https://markets.businessinsider.com/news/currencies/chinas-ministry-of-public-security-launches-blockchain-based-real-name-decentralized-identifier-system-1032891627.

[56] Biometric Update, "China's Project to Verify Real-Name Digital ID Leans into National Blockchain Ambitions," December 13, 2023. https://www.biometricupdate.com/202312/chinas-project-to-verify-real-name-digital-id-leans-into-national-blockchain-ambitions.

[57] Huang, R., "A 2024 Overview of the e-CNY: China's Digital Yuan," Forbes, July 15, 2024. https://www.forbes.com/sites/digital-assets/2024/07/15/a-2024-overview-of-the-e-cny-chinas-digital-yuan/.

[58] DID Alliance, "DID Alliance website," 2024. https://www.didalliance.org/.

[59] Jamaica Information Service, "$2,500 Incentive for Jamaicans to Get Digital Wallet," March 10, 2022. https://jis.gov.jm/2500-incentive-for-jamaicans-to-get-digital-wallet/?fbclid=IwAR2Cjjnrm0p215AdrDRpalPF-V9jEUj3KagONj3HwN55jLK3PwbFMv7wfgU.

Chapter 4

On the Economics of Cybersecurity Breach Reporting

Carlos Hernandez

Data breaches represent a permanent threat to all types of organizations. Although the type of breaches differ, the impacts are always the same. The financial ramifications of these breaches are significant, leading to stock price volatility, legal challenges, and reputational damage. For businesses, it may seem advantageous to avoid reporting breaches in order to reduce some of these consequences. However, withholding information about compromised data increases the risks of identity theft, financial fraud, and other forms of harm to the affected individuals. This creates a conflict between companies and their customers regarding how breaches should be handled. To address this, data breach notification laws were introduced to ensure that companies inform affected parties when their data is compromised. While the specifics of these laws vary by location, most require companies to notify victims within a certain time frame. As a result, these laws limit companies' flexibility and autonomy in managing such incidents. The effectiveness of data breach notification (DBN) laws in protecting individuals from more severe damage remains a point of debate. While these laws aim to safeguard consumers by ensuring they are informed when their data is compromised, they may also impose significant hardships on companies. These regulations can create additional burdens for businesses without necessarily improving the overall situation for customers.

Contents

4.1 Introduction

This section of the paper will provide a brief overview of the problem at hand and the possible motivations behind the solutions.

4.1.1 Problem

The frequency and financial impact of cybercrime have grown significantly in recent years. In 2020, 791,790 cybercrime incidents were reported, resulting in over $4.1 billion in losses [10]. By 2023, these numbers increased to 880,418 incidents with potential losses exceeding $12.5 billion [11], indicating not only a rise in cases but also a disproportionate escalation in associated costs.

By Complaint Loss	
Crime Type	*Loss*
Data Breach	$534,397,222
Phishing/Spoofing	$18,728,550

Figure 4.1: Losses caused by data breaches compared to losses caused by phishing/spoofing cases in the year 2023 [11].

Data breaches, though accounting for only 3,727 of the incidents in 2023, caused $534,397,222 in losses—far exceeding the $18,728,550 caused by the most common cybercrime type, phishing/spoofing, despite its 298,878 reported cases [11]. To mitigate such significant damages, governments have implemented data breach notification (DBN) laws, which require firms to disclose breaches to affected parties. These laws have influenced corporate behavior, particularly in cybersecurity practices and incident management. However, the economic implications of these laws—both for firms and society—remain underexplored, creating a critical need for further study.

4.1.2 Motivation

The increasing frequency and costs of cybercrime pose significant threats to businesses, consumers, and the broader economy. Among various forms of cybercrime, data breaches stand out because of their disproportionately high financial impact relative to their frequency. These breaches not only compromise sensitive data, but also erode consumer trust and impose substantial recovery and legal costs on companies.

To address this growing concern, governments worldwide have introduced data DBN laws. These regulations aim to improve transparency by requiring companies to inform affected individuals and authorities when breaches occur. Although DBN laws have undoubtedly influenced corporate behavior, their broader economic implications remain not explored sufficiently. For example, questions arise about how these laws affect firms' incentives to invest in cybersecurity, the timing and manner of breach disclosures, and the associated costs for businesses and society.

Understanding the economic dynamics of DBN laws is critical for shaping effective policy frameworks that balance consumer protection with corporate sustainability. This study seeks to contribute to this understanding by examining how DBN laws influence the costs and management of cybersecurity incidents, the market reactions to breach disclosures, and the long-term incentives for firms to adopt robust cybersecurity measures.

4.2 Economics of Cybersecurity Breach Reporting

This part of the text will explain the nature of DBN laws and highlight the differences between these laws across different countries.

4.2.1 Notification Laws

Notification laws aim to address the increasing demand from customers for stronger protection against data breaches, which present severe risks to the security and privacy of sensitive information. These breaches often result in the exposure of personal data, which cybercriminals can exploit for malicious purposes, such as payment fraud, identity theft, and phishing schemes. Once compromised, this information can be sold on the dark web, used to open unauthorized accounts, or even used in elaborate scams that can devastate victims financially and emotionally.

The repercussions of a data breach extend far beyond monetary losses. They significantly impact the trust and loyalty customers place in a company. When clients trust businesses with their sensitive information, such as payment details, contact information, or social security numbers, they expect it to be securely stored and protected. A breach of this trust not only damages the company's reputation but also erodes the relationship it has built with its customers. For many businesses, especially those in highly competitive industries, this loss of trust can result in decreased customer retention, negative publicity, and a lasting hit to brand value [8].

One of the most troubling aspects of data breaches is how they are often invisible to customers. Many breaches are discovered long after they occur, and the individuals affected may remain completely unaware until months—or even years—later [8]. This delayed notification means that by the time customers are informed, the damage may already be done. For example, their financial accounts could be compromised, their credit ratings affected, or their personal details used in identity theft schemes. The lack of timely awareness leaves customers unable to take immediate steps to mitigate the harm, such as freezing credit, changing passwords, or monitoring financial transactions.

Notification laws play a critical role in closing this gap. By mandating that companies inform affected parties of data breaches promptly, these laws aim to empower customers to take control of their security and minimize potential damage. Furthermore, such laws act as a catalyst for companies to enhance their cybersecurity measures proactively. Knowing they are required to disclose breaches increases the incentive for businesses to invest in better data protection technologies and protocols. In doing so, notification laws not only help individual customers but also contribute to creating a safer digital environment overall.

Data breach notification (DBN) laws serve as a crucial mechanism to mitigate these risks. By mandating that companies disclose breaches to affected parties, these laws aim to create transparency and accountability, incentivizing businesses to invest more heavily in cybersecurity infrastructure. This, in turn, reduces the likelihood of future data infringements and reassures customers that they will be informed if a breach occurs. Enhanced transparency builds consumer confidence, as users are assured that their rights and data are taken seriously. From this perspective, one might assume that it is always in a company's best interest to inform users about data breaches promptly. However, the reality is far more complex. Companies often weigh the benefits of disclosure against potential reputational and financial costs, which can lead to reluctance in reporting breaches unless legally required.

The introduction of DBN laws began in California in 2002, marking a significant turning point in data privacy regulations [15]. California's pioneering legislation required companies to notify affected parties when their data was compromised. Specifically, the

law stated that users must be informed if "unencrypted personal information was, or is reasonably believed to have been, acquired by an unauthorized person" [2]. This legislation set a precedent, and over time, all other U.S. states implemented similar laws, with Alabama being the last to do so in 2018. Despite the lack of uniformity among state laws–since requirements and enforcement vary–these state-level DBN laws share a common objective: to promote public awareness and enhance personal internet security. Research has shown that states introducing DBN laws experienced a reduction in annual identity theft cases, indicating the effectiveness of these measures in addressing cyber risks.

In Europe, the General Data Protection Regulation (GDPR) has introduced a more stringent and comprehensive framework for data protection. The GDPR applies to companies that handle the data of EU citizens, regardless of where the company is located. Regarding data breach notifications, the GDPR mandates that companies disclose breaches within a strict 72-hour window after detection. Failure to comply with this timeline can result in severe legal and financial repercussions. Since its enforcement in 2018, the GDPR has become the most robust data protection regulation globally, setting a high standard for corporate accountability. The law has had significant financial implications, costing companies billions in compliance expenses and fines. Major corporations such as Google, Amazon, and Meta have faced substantial penalties under the GDPR, reinforcing its role in fostering a more secure digital environment for users. These fines serve as a powerful deterrent, compelling companies to prioritize data security and transparency.

Switzerland, while outside the jurisdiction of the EU, has implemented its own data protection framework under the Federal Act on Data Protection (FADP). The FADP requires companies to notify the Federal Data Protection and Information Commissioner (FDPIC) in the event of a data breach [5]. Unlike the GDPR, the FADP does not impose a fixed notification timeframe, stating only that breaches must be reported "as soon as possible." This flexibility provides companies with more leeway in managing breach disclosures but also leaves room for potential delays in notifying affected parties. While this approach reduces the immediate pressure on companies, it may also weaken the overall effectiveness of the regulation compared to stricter frameworks like the GDPR.

4.2.2 Unexpected Cost of Breaches

Data breaches are profoundly damaging to companies, both financially and reputationally. While this paper has discussed the approximate monetary losses they cause, other significant impacts remain understated. Among these is the reputational damage that often accompanies a breach. Companies depend heavily on customer trust to grow and maintain a sizable client base. This trust hinges on the confidence customers have in the company's ability to securely handle their personal data while delivering the desired service. When a data breach occurs, it exposes vulnerabilities in the company's internal controls and security measures. This disclosure can lead customers to question the company's reliability, making them hesitant to continue their business relationships [3].

Rebuilding trust after a data breach or security incident is not easy and can take a long time. It requires companies to spend a lot of time, money, and effort to fix the damage and show customers they can be trusted again. During this process, businesses often have to pause or delay other important projects, which can hurt their overall productivity.

One of the first things a company needs to do is improve its cybersecurity systems. This might mean updating security software, hiring experts to find and fix weaknesses, or putting in place better ways to protect customer data. These upgrades are necessary to prevent another breach, but they can be very expensive, especially for smaller companies. At the same time, companies need to work on their public image. They often release statements apologizing for the breach and explaining what they are doing to fix the

problem. Some businesses also hire public relations experts to help them communicate better with the public. This is important for showing that they take the issue seriously, but it adds to the costs [3].

Another important step is helping customers who were affected. This could include paying for credit monitoring services, offering refunds, or providing other types of support to reduce the impact on their customers. While this shows the company is taking responsibility, it can be very costly. Internally, companies have to train their employees on new security measures and spend a lot of time managing the crisis. This can slow down regular work, delay new products, or reduce sales because employees are focused on fixing the problem instead of their normal tasks. To top it all off, the long-term effects can be hard to recover from. Customers may stop trusting the company and switch to competitors, sales might drop, and the company's reputation could take years to rebuild. For businesses in highly competitive industries, losing customer trust can be especially damaging [15].

Beyond reputational damage, data breaches have a destabilizing effect on a company's financial stability, particularly its stock price. Following the disclosure of a breach, stock prices often experience a sharp decline as investor confidence wavers. This decline amplifies the financial strain on the company, as any missteps in managing the aftermath of the breach can lead to further losses. The increased risk of a stock price crash in the wake of a breach announcement puts firms in a precarious position, making it harder to secure funding or bank loans. Financial institutions may become wary of the company's reliability, further complicating efforts to navigate this turbulent period.

To mitigate reputational damage and avoid market overreactions, managers may resort to withholding or downplaying other bad news within their control. This behavior stems from a fear of exacerbating the firm's already fragile standing. Companies in industries prone to cyberattacks may face similar reputational challenges, even without experiencing a breach themselves. Negative perceptions may arise merely by association if other companies within the same industry are targeted. Consequently, managers in such firms may feel pressured to manipulate market perceptions, including concealing unfavorable news, to protect their company's reputation and stability.

Moreover, data breach disclosures have been linked to opportunistic behavior among corporate insiders. Research has shown that insider selling tends to increase when companies are required to reveal a breach [4]. This opportunistic selling can erode investor confidence further, adding another layer of complexity to an already challenging situation. Such behavior also incentivizes managers to adopt a shortsighted approach, potentially withholding or delaying the release of other negative information to avoid compounding the market's reaction [15].

4.2.3 Impact on Companies

To illustrate how devastating a data breach can be, we can take the Equifax data breach as an example. Equifax, a credit reporting agency, maintains comprehensive databases of consumer and business information. Information like date of birth, social security number, names and credit account information of their customers. On the time of the attack Equifax had over 8500 vulnerabilities that they had failed to address, one of these vulnerabilities was used by the hacker to intercept the systems. This maneuver led to the information of 147 million Americans being compromised, but the worst part of this incident was Equifax's decision to not announce the hack to the public until 6 weeks after the incident. This poor management of the situation resulted in Equifax having to pay $700 million in fines and compensations [1].

A data breach itself cost companies overall an average amount of 9.44 million in 2022 [3]. So if data breaches help companies secure their systems it would be logical to assume

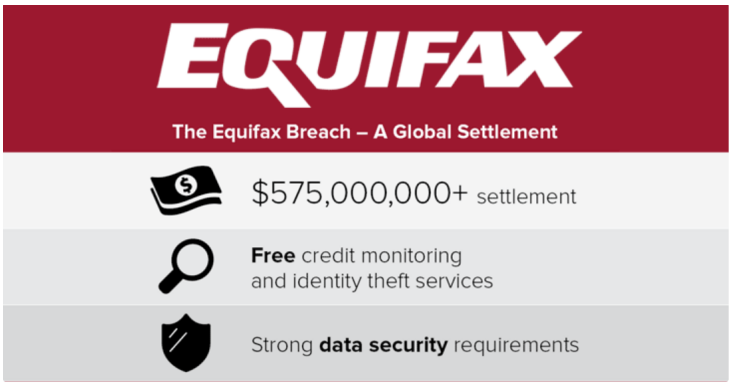


Figure 4.2: Consequences of the equifax data breach, out of the $700 million they had to pay $575 million went into settlements [17].

that companies would prioritize measure that incentivize decisions that encourage them to better their security, meaning that reporting the attack would be an obvious choice, but the reality differs. The reporting of the data breaches itself has some consequences in a company's finances and reputation. Corporate managers are often motivated to hoard the bad news about data breaches to navigate the effect of the disclosures with less risk [3]. Whenever data breaches happen the person detecting it, might be seen as the responsible for this incident, meaning that he will be seen as the one who did not do a good enough job securing the systems, and might be the one having to carry all of the burden. So it is important to remember it's in a managers best interest to withhold negative information from investors to protect their job security, compensation and reputation [3]. Leading to firms underreporting cyber incidents as much as possible.

This is where DBN laws come into play as they mandate companies to report cyber-incidents it removes the flexibility of the decision making for the managers. Having this in mind companies have adapted their strategies to utilize the exemptions and other loopholes in these laws. In the United states for example companies have taken measures to make their data more secure by investing more heavily in encrypting their data in states which have implemented data breach notification laws, because these laws exempt the breach of encrypted data[8]. Since the introduction of DBN laws in California for example, the first state to do so, companies were more encouraged to update to newer versions of their server software. Companies that used apache servers for example there was a surge of 1.7 - 2.7 percent to keep their servers more updated, while larger firms decreased the technological age of their web server by to 2 up to 7.8 percent. So while the first law had an impact on the safety of the servers it wasn't specially significant[14].

Companies that decide to not comply with the regulations must be prepared to face consequences. The punishment for non-compliance in the GDPR are, depending on the severity of the case, can go up to 2% of the global annual revenue of the company or a fine of 10 million dollar (depending on which one of these numbers are higher) for less severe cases. For the more severe cases the punishment can go up to 4% of the global revenue or a $20 million fine, the same logic applies as with the less severe cases. UBER for example suffered a data breach in 2018, instead of reporting it after its detection UBER decided to pay off the responsible hacker, so he could fix the bug and made him sign an NDA to not talk about this incident publicly. To their dismay the truth still came to light which lead them to be investigated and finally having to settle for $ 148 million[9].

4.3 Summary

This paper examines the intricate relationship between data breaches, corporate behavior, and the economic implications of data breach notification (DBN) laws. It highlights the

increasing frequency and financial impact of cybercrime, illustrating how data breaches, though less frequent than other cybercrimes, cause disproportionately higher losses and reputational damage to companies.

DBN laws were introduced to enhance transparency and incentivize companies to improve their cybersecurity measures. These laws mandate timely disclosure of breaches to affected parties and regulators, aiming to mitigate harm and bolster public trust. The paper outlines key differences in regulatory approaches, such as the GDPR in Europe, which imposes strict timelines and severe penalties for non-compliance, and the more flexible FADP in Switzerland. The analysis demonstrates how these frameworks have prompted companies to adapt strategies, such as encrypting data and updating server software, while also revealing loopholes and opportunities for evasion.

The paper concerns itself with the broader consequences of data breaches for companies. Beyond immediate financial losses, breaches erode customer trust, destabilize stock prices, and complicate access to funding. Managers often face conflicting incentives, balancing legal requirements with reputational risks and personal interests, such as insider trading opportunities or preserving job security. The underreporting of breaches remains a critical challenge, driven by these competing priorities. Despite their intent, DBN laws have unintended consequences, including the potential for firms to hoard negative information or strategically manipulate market perceptions.

However, compliance remains essential, as demonstrated by cases like Uber, where attempts to conceal a breach led to severe penalties and reputational harm. Ultimately, the study underscores the importance of a balanced regulatory framework that not only enforces transparency but also supports companies in their cybersecurity efforts. It concludes that while DBN laws are a step toward a more secure digital environment, continuous evaluation and refinement are necessary to address their limitations and maximize their positive impact on businesses and society.

Appendix: Definition of Key Terms

To have a clear overview of the topics that are going to be discussed, there are some concepts we first have to introduce to establish an understandable vocabulary for this paper.

A data breach is defined as an electronically mediated service failure that occurs when sensitive financial, personal, or customer data is released to or accessed by parties external to the organization. It tarnishes a company's reputation and destabilizes its relationship with customers [13]. Data breach notification laws: They serve important purposes; they provide an incentive for organizations to protect sensitive data and to actually inform users that their data has been compromised [8].

Bibliography

[1] Berghel, H. (2019). The Equifax Hack Revisited and Repurposed.

[2] California Legislative Information (2024) https://leginfo.legislature.ca.gov/faces/codes_displaySection.xhtml?lawCode=CIV&sectionNum=1798.82

[3] Cao, H., Phan, H. V., & Silveri, S. (2024). Data breach disclosures and stock price crash risk: Evidence from data breach notification laws. *International Review of Financial Analysis*, 93, 103164-. https://doi.org/10.1016/j.irfa.2024.103164

[4] Chen, X., Hilary, G., & Tian, X. S. (2019). Data breach disclosure and insider trading [Working paper].

[5] Fedlex the publication for federal law. Federal Act on Data Protection (2023) https://www.fedlex.admin.ch/eli/cc/2022/491/en

[6] Romanosky, S., Telang, R., & Acquisti, A. (2011). Do data breach disclosure laws reduce identity theft? *Journal of Policy Analysis and Management*, 30(2), 256—286.

[7] Ruohonen, J., & Hjerppe, K. (2022). The GDPR enforcement fines at glance. *Information Sciences*, 106, 101876.

[8] Sullivan, R., & Maniff, J. (2016). Data Breach Notification Laws.

[9] Conger, K. (2018). UBER settles data breach investigation for $148 million. *New York Times*. https://www.nytimes.com/2018/09/26/technology/uber-data-breach.html

[10] Federal Bureau of Investigation (FBI). (2020). Internet Crime Report 2020. Available at https://www.ic3.gov/AnnualReport/Reports/2020_IC3Report.pdf.

[11] Federal Bureau of Investigation (FBI). (2023). Internet Crime Report 2023. Available at https://www.ic3.gov/AnnualReport/Reports/2023_IC3Report.pdf.

[12] General Data Protection Regulation (GDPR) (2024) https://gdpr.eu/tag/gdpr/

[13] Hammouchi, H., Cherqi, O., Mezzour, G., Ghogho, M., & Koutbi, M. E. (2019). Digging Deeper into Data Breaches: An Exploratory Data Analysis of Hacking Breaches Over Time. *Procedia Computer Science*, 151, 1004—1009. https://doi.org/10.1016/j.procs.2019.04.141

[14] Muricano-Goroff, R. (2019). Do Data Breach Disclosure Laws Increase Firms' Investment in Securing Their Digital Infrastructure? *Boston University*. https://weis2019.econinfosec.org/wp-content/uploads/sites/6/2019/05/WEIS_2019_paper_33.pdf

[15] Obaydin, I., Xu, L., & Zurbruegg, R. (2024). The unintended cost of data breach notification laws: Evidence from managerial bad news hoarding. *Journal of Business Finance & Accounting*, 51(9—10), 2709—2736. https://doi.org/10.1111/jbfa.12794

[16] Skinner, D. J. (1994). Why firms voluntarily disclose bad news. *Journal of Accounting Research*, 32(1), 38—60.

[17] The Hacker News: Equifax to Pay up to $700 Million in 2017 Data Breach Settlement `https://thehackernews.com/2019/07/equifax-data-breach-fine.html`

Chapter 5

Impact of Data Localization Laws on Global Trade and Economics

Daniel Ritter

Data localization laws, which mandate the storage and processing of data within specific geographical boundaries, have become increasingly prevalent in recent years. In a digital age in which 153.52 zettabytes of data will be created in 2024 alone [1], it is necessary to protect personal data. Governments have enacted regulations to give natural persons more rights in the digital realm and to prevent data breaches that have occurred in the past [2]. This paper dives into the European act of the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA) and examines the potential impact of these laws on global trade and economic growth. It analyzes the effects on global business competitiveness and digital innovation. This work presents a technical approach to implement these regulations through encryption. By evaluating the costs and benefits of data localization, this document aims to contribute to a more informed understanding of its implications for consumers and businesses alike.

Contents

5.1 Introduction and Problem Statement

In the digital age, the need for regulation and data protection is growing and becoming essential for consumers and businesses. In recent years, governments have developed laws to protect the rights of their citizens and introduced more regulations for companies to collect, store, share, and retain data. A few years ago, the European Union implemented the General Data Protection Regulation (GDPR) and soon after, in the United States, California passed its own data privacy law, the CCPA [25]. In Switzerland, the new Federal Act on Data Protection (nFADP) was implemented and companies were required to comply with this legislation from September 1, 2023 [3]. Efforts were also made to introduce regulations in other states or countries [23]. These laws are intended to protect personal data. Although the impact for end users is particularly that they will have greater rights and protection of their data, companies face challenges in implementing these regulations. These come with costs and impact the company growth and digital innovation [22]. For example, IoT companies face significant financial burdens when complying with data protection regulations. Some estimates suggest that compliance costs for companies in the IoT segment could increase on average three to four times, in certain cases even more by eighteen times, compared to previous regulatory standards [27] [28]. Many companies are concerned: In 2017, almost half of firms globally feared they won't meet the fast approaching regulatory deadline for the 99 articles of the General Data Protection Regulation (GDPR) [32] that became enacted on May 25, 2018 [36].

5.2 Legal Background

5.2.1 General Data Protection Regulation (GDPR)

The General Data Protection Regulation (GDPR), a modern data protection framework, superseded the Data Protection Directive (Directive 95/46/EC) of the European Union which was passed in 1995. The Directive, outdated for the digital age, prompted the development of the GDPR [36]. The GDPR was proposed by the European Commission on January 25, 2012 [30]. It was adopted by the European Parliament and Council on April 14, 2016, and became enacted on May 25, 2018 [36]. Although the Directive sets out specific goals, each member state has the flexibility to determine how to implement the Directive into their national laws. In contrast, regulations are directly applicable in all member states and become effective on the date specified by the European Union [27]. The primary objective of the GDPR is to empower individuals with greater control over their personal data [35]. The GDPR imposes stringent regulations on organisations that collect and process personal data, requiring them to adhere to strict security measures and obtain explicit consent for data usage [10]. The regulations ensure the free flow of personal data between EU member states, but not outside of these borders [27].

5.2.1.1 Definitions and Entities

Personal data The GDPR defines *personal data* broadly to include any information that can identify a person, either directly or indirectly. This includes names, dates of birth, location data, IP addresses, cookie identifiers, and more. The GDPR also has a special category for "sensitive personal data" which includes genetic data, biometric data, and health-related information [5] [8].

Data subject A *data subject* is any identified or identifiable natural person that can be identified directly or indirectly by any unique characteristics [5] [18].

Data controller A *data controller* is an individual or organisation who holds personal data and determines the purposes and means of processing personal data [37]. The data controller is responsible for making decisions about how data is collected, used, shared and protected [36].

Data processor A *data processor* is responsible for processing personal data on behalf of a data controller [37]. They carry out the instructions of the data controller, but do not have the primary responsibility for determining the purposes of processing [36]. The GDPR has special rules for these individuals and organisations [7].

Data processing *Data processing* is any action performed on data, whether automated or manual. This includes collecting, recording, organizing, structuring, storing, using and erasing of data [7].

Data Protection Officer (DPO) Under the GDPR, data controllers are required to designate a *Data Protection Officer* (DPO) and make their contact details available to the public if the organisation is a public authority, the core activities are to monitor people systematically, the organisation processes data in "large volumes" or the collected data is considered "sensitive data" [7]. The DPO can be hired as an independent contractor or can be an employee of the controller [11] [34].

Data Protection Authorities (DPAs) *Data Protection Authorities* are independent public authorities that supervise the enforcement of data protection laws in their respective jurisdictions.

Data breach A *data breach* occurs when unauthorized individuals or entities gain access to sensitive or confidential information. Data breaches can have serious consequences for companies, including financial loss, reputation damage, or legal penalties and for end users, financial loss, breach and open access of personal data [5].

5.2.1.2 Consumer Rights

The GDPR establishes data rights for EU residents, such are [6] [36]:

- **Right to be informed**
  Individuals have the right to be informed about the processing of their personal data.
- **Right of access**
  Individuals have the right to access their personal data.
- **Right to rectification**
  Individuals have the right to have inaccurate or incomplete personal data corrected.
- **Right to erasure**
  Individuals have the right to request the erasure of their personal data. This is also known under the right: *Right to be forgotten* [9].
- **Right to restrict processing**
  Individuals have the right to request a restriction on the processing of their personal data.

- **Right to data portability**
  Individuals are entitled to obtain their personal data in a structured, commonly used, and machine-readable format, and have the right to transfer it to another controller.

5.2.2 California Consumer Privacy Act (CCPA)

The California Consumer Privacy Act (CCPA) is a modified version of the Californians for Consumer Privacy (CCP) initiative that was signed into law on June 28, 2018 [33]. The primary objective of the CCPA is to enhance the privacy rights of California residents. The CCPA was amended by California voters in November 2020 and the California Privacy Rights Act (CPRA) [14] went into effect in January 2023 [13] [33]. For the sake of simplicity, in the following both acts are referred to as CCPA.

5.2.2.1 Definitions and Entities

Personal information The CCPA defines *personal information* as any information that identifies, describes, or could reasonably be linked, directly or indirectly, by a particular consumer, household or device [15]. This includes names, addresses, IP addresses, and other identifiers, but does not cover the same breadth of "sensitive data" as the GDPR. For example, medical information is exempt from CCPA [15].

Business The California Consumer Privacy Act (CCPA) defines a *business* as any legal entity, regardless of location, that operates for profit and engages in activities that constitute "doing business" in California and collects personal information from California consumers [14] [15].

Consumer A *consumer* is a natural person who resides in California. This includes individuals who are temporarily outside the state but are still legally California residents [14].

5.2.2.2 Consumer Rights

The CCPA aims to give consumers more control over their personal information by granting them specific rights [13] [14], such are:

- **Right to know**
  Consumers have the right to request disclosure of the personal information collected about them.
- **Right to delete**
  Consumers can request that businesses delete the personal information collected from them.
- **Right to opt-out**
  Consumers have the right to opt-out of the sale or sharing of their personal information.
- **Right to opt-in**
  Consumers under the age of 16 have the right to opt-in to the sale of personal information.
- **Right to non-discrimination**
  Businesses are prohibited from discriminating against consumers for exercising their CCPA rights.

- **Right to initiate a private cause**
  Consumers have the right to initiate a private cause of action for data breaches.

- **Right to correct**
  Introduced by the CPRA, consumers have the right to correct inaccurate personal information that a business has about them.

- **Right to limit**
  Under the CPRA amendments, consumers have the right to limit the use and disclosure of their sensitive personal information for unauthorized purposes.

5.2.3 Similarities and Differences

The GDPR applies to all organisations, including businesses, public bodies, and non-profit institutions, that process personal data of EU/EEA residents, regardless of the organisation's location [4] [12]. This includes all 27 EU countries, as well as Iceland, Liechtenstein, and Norway. In contrast, the CCPA is limited to for-profit businesses that operate in California and meet specific criteria, such as having an annual gross revenue of more than $25 million, collecting, buying, or sharing the personal information of 50,000 or more Californian consumers, or deriving more than 50 % of their annual revenue from the sale of personal information [15] [16].

Under the GDPR, businesses must have a lawful basis to collect and process personal data [15]. There are six of them. If *Consent* is the basis, it must be explicit and affirmative from the data subjects before data collection. This includes an opt-in requirement for cookies that track personal data [15]. If *Contract* is the basis, the data processing is necessary to fulfill a contract (e. g. delivering a product or service) with the person, or to take steps before entering a contract [15]. If a company needs to use the data to comply with a law or regulation we have a *Legal obligation* [15]. We have *Vital interests* when the processing is necessary to protect someone's life, safety, or well-being [15]. The basis is a *Public task* if an organisation needs the data to perform a task with a clear legal basis that is in the public interest, e.g. by government or law enforcement [15]. And there is a *Legitimate interest* when a company has a legitimate business interest that requires processing personal data, e.g. an insurance company processing data to prevent fraud that may affect customers [15]. Companies must be able to justify which of these legal grounds they are relying on each time they use personal data. If consent is the legal basis, organisations must also be able to demonstrate that consent was obtained and also demonstrate that it was obtained in a valid manner, i.e. that the consent was freely given, specific, informed and unambiguous [15]. The CCPA does not require a lawful basis for processing personal information. Instead, businesses can process personal information for any purpose unless the consumer exercises their right to opt-out. Opt-in is only mandatory for consumers under the age of 16 [16].

Under the GDPR, the age of consent is 16, although member states can lower it to 13 if they choose. Parental consent is mandatory for those below the age of consent [7] [16]. Under the CCPA, parental consent is mandatory for consumers below 13 years old. For those 16 and above, consent is not mandatory and opt-out is the primary mechanism [16].

The GDPR requires organisations to implement appropriate security measures based on the risk involved in processing personal data [7] [16]. The CCPA does not specify particular security requirements but imposes a privacy right of action against businesses for inappropriate security measures [15] [16].

Under the GDPR, fines for non-compliance can be up to 10 million euro or 2 % of annual global turnover for less severe violations and up to 20 million euro or 4 % of annual global turnover for severe violations, whichever is higher [7] [18] [31]. These fines are imposed by member state data protection authorities (DPAs) [16]. Under the CCPA,

fines can be up to $2,500 for each violation and $7,500 for each intentional violation. Additionally, consumers can receive $100 to $750 per consumer affected in a breach [15]. These fines are imposed by state courts [16].

Similarities and Differences	**CCPA**	**GDPR**
Effective Date	January 1, 2020	May 25, 2018
Scope	Applies to for-profit businesses that hold personal information of California residents and meet some conditions.	Applies to businesses that hold personal data of EU/EEA residents.
Personal Data	Information that relates to an individual, household or device. Excludes publicly available personal information recorded by federal, state or local government.	Data that relates to a living individual used for commercial purposes. Excludes publicly available information.
Opt-in necessary for Data Collection	No, unless the consumer is under 16 years old.	Yes
Right to Opt-out	Yes	Yes
Age of Consent	16 and below. Parental consent mandatory for consumers below 13 years.	16 (Member State laws can lower it to 13). Parental consent mandatory for those below 16.
Legal Basis for Data Processing	No specific legal basis but it provides exceptions and allows data use for business purposes.	Six legal reasons for data use: *Consent, Contract, Legal obligation, Vital interests, Public task, Legitimate interest*
Data Security	No specific security requirements but businesses face legal action for inappropriate security measures.	Requires organisations to implement appropriate security measures according to the risk involved.
Fine	Up to 4 % of global revenue or 20 million euro, whichever is greatest.	Up to $7,500 per violation, private right of action for data breaches.

Table 5.1: Similarities and Differences of CCPA and GDPR [15] [16]

5.3 Potentials and Implications

5.3.1 Impact on Consumers

When focusing on consumers, there are several impacts worth discussing. One advantage of data protection laws is that they prioritize the security and protection of individuals' personal data. Another positive effect is that users gain greater control over their personal data, including decisions about what information is shared and with whom.

On the other side, the abundance of consent options can be overwhelming and confusing for users. Businesses may pass on increased operating costs to consumers, potentially raising prices for services, especially subscription-based ones. Global service availability may be restricted due to varying compliance requirements across regions [22]. Localized services tailored to comply with regional laws might lead to inconsistent experiences across countries. For instance, digital online services like Facebook could differ significantly from one country to another.

These laws can also act as protectionist measures, potentially limiting the advantages of a globally connected Internet [22]. Consumers might miss out on a wider range of products, services, and global commerce opportunities [22]. Splitting the global Internet into regional systems can hinder the efficiency and innovation that stems from global data sharing [22]. This may result in fewer new products and services being developed and made available to consumers [22]. For example, the direct welfare loss caused by the GDPR is estimated at approximately 260 euros per European citizen [29].

5.3.2 Impact on Businesses

Implementing compliance measures can help decrease the risk of data breaches for companies [39]. By enforcing stricter data protection practices, for example encryption and crypto-shredding, organisations can better safeguard sensitive information [39].

Although there is no definitive evidence of a direct relationship between GDPR compliance and increased customer trust [28], some businesses may experience improved reputation and customer confidence by demonstrating commitment to data protection.

Non-compliance with GDPR can result in hefty fines. Since the GDPR came into force, more than 1,000 fines have been imposed, with most cases targeting small and medium-sized businesses [21]. Notable examples of high-profile fines include:

- In May 2023, the Irish Data Protection Commission (DPC) imposed a historic fine of 1.2 billion euro on US tech giant Meta [19].
- On July 16, 2021, the Luxembourg National Commission for Data Protection (CNDP) issued a fine in the amount of 746 million euro to Amazon.com Inc [19].
- Due to violations of GDPR, with a specific focus on its handling of children's accounts, TikTok faced a substantial fine of 345 million euro [19].
- The Dutch Data Protection Authority (DPA) has fined Uber 290 million euro for unlawfully transferring personal data of European taxi drivers to the US [19].
- On September 2, 2021, the Irish Data Protection Commission (DPC) announced its decision to impose a GDPR fine on WhatsApp of around 225 million euro following a three-year investigation [19].
- On December 31, 2021, the Commission Nationale de l'Informatique et des Libertés (CNIL) fined Google LLC 90 million euro for not allowing YouTube users in France to reject cookies as easily as they could accept them [19].

These substantial penalties underscore the importance of adhering to regulations. 86 % of firms think non-compliance will have a major negative impact on their business [32].

But achieving GDPR compliance can be extremely challenging and expensive for businesses, potentially costing millions of euros [29]. An infographic from Veritas Tech LLC, a leader in enterprise data management, shows a substantial investment required to deliver on GDPR readiness journey. On average, firms are forecasting spending in excess of 1.3 million euro on GDPR compliance [32]. A third of respondents fear their current technology stack is unable to manage their data effectively that is hindering GDPR compliance [32]. 39 % of respondents say their organisation cannot accurately identify and locate relevant data [32]. And 42 % admit to having no system in place to determine which data should be saved or deleted based on its value [32].

The GDPR mandates that data controllers must be able to demonstrate their compliance with its regulations [7]. One possible way to do so is to train stuff and implement technical and organisational security measures [7]. To obtain for example a GDPR certificate, the total GDPR compliance fee can range from $20,500 to $102,500 depending on the size and complexity of the organisation [20]. To comply with GDPR, companies must also obtain ISO 27701 and ISO 27001 certifications, which is roughly estimated to incur an additional implementation cost of approximately $15,000 to $70,000 [20] [24].

Implementing compliant systems and processes can add layers of complexity to existing infrastructure. This increased complexity may inadvertently create new security vulnerabilities if not managed carefully. There have already been unintended creation of new cyber risks due to GDPR compliance measures [28]. For example, while the GDPR and the CCPA are designed to empower users to control their data through facilitated user requests, they inadvertently provide opportunities for hackers and identity thieves due to the lack of user authentication. As a result, companies are required to develop data pools to handle these requests, which creates an attractive target for cybercriminals [29].

And then, low compliance rates among eligible firms, with many small business owners are unaware of GDPR requirements and potential fines [28]. “Indeed, less than half of eligible firms are fully compliant with the GDPR; one-fifth say that full compliance is impossible. In a recent survey of small business owners in the EU, a whopping nine out of ten reported not knowing about the GDPR and that its fines for non-compliance could adversely impact them” [28].

5.3.3 Impact on Global Trade and Economics

The potential and implications of data localization laws on global trade and economics are diverse. On one side, data localization laws prioritize and protect the privacy rights of consumers, ensuring their personal data is handled responsibly. And these laws could lead to increased consumer trust and improved quality in digital products and services worldwide but doesn't have to. Lack of evidence showing increased consumer trust in the digital ecosystem despite GDPR-type regulations [28]. Likewise, California has more privacy laws than any other state and yet residents do not report feeling more private or secure [29].

On the other side, data localization requirements can hinder international trade by restricting cross-border data flows [22]. That is crucial because “globally, half of all services trade depends on access to cross-border data flows” [22]. Likewise, 12 % of global goods trade occurs online [22]. The forcement of companies to fragment their data storage can lead to inefficiencies and potentially increased cybersecurity risks [29]. The restricting of cross-border data flows can result in higher operational costs for businesses operating globally. These high operational costs are passed on to prices and the end user is ultimately the one who suffers. Data localization laws can create barriers in the global digital economy, potentially slowing down technological advancements and economic growth on

a global scale [22]. Countries implementing strict data protection laws may become less appealing to foreign investors due to increased complexity and compliance costs [22]. The restrictive nature of data localization laws could decrease trade volumes and efficiency, potentially leading to reduced economic growth, higher prices and increased poverty.

High compliance burdens may lead to decreased competition in the global market, potentially favoring large tech companies and creating a less free market environment [28] [29]. High compliance costs (approximately $3 million for an average firm of 500 employees in 2019) leading to market exits [28] [29]. In 2017, 18 % of businesses globally thought, the high penalties could cause them to go out of business [32].

Research indicates a decrease in EU website traffic following GDPR implementation [26] [36]. There has been a weakening of small and medium-sized enterprises (SMEs), with some ad tech competitors losing up to one-third of their market position [28] [29]. There has also been a withdrawal of many US media, retailers, game companies, and service providers from the EU market [28]. Over 1,000 US newspapers no longer show their content in the EU due to compliance concerns [28] [29]. In Northern Europe, the use of digital platforms for children has become increasingly difficult due to parental consent requirements for users under the age of 13 [28] [29].

WHOIS information obscurity leading to potential cybersecurity risks [28]. "The Internet Corporation for Assigned Names and Numbers (ICANN) recently announced a Temporary Specification [38] that allows registries and registrars to obscure WHOIS information they were previously required to make public, ostensibly to comply with the GDPR" [28].

The GDPR poses challenges for innovation and research. Many of its requirements are fundamentally at odds with big data, artificial intelligence, blockchain and especially machine learning. Particularly problematic are provisions that require data processors to disclose the purpose of data processing, minimize data use, and automate decision-making [29].

5.4 Possible Technical Approach

5.4.0.1 Right to erasure / be forgotten

The General Data Protection Regulation (GDPR) includes a crucial provision known as the "Right to erasure" or "Right to be forgotten". This is detailed in Chapter 3, Article 17, Point 1 of the GDPR [9], which states:

"The data subject shall have the right to obtain from the controller the erasure of personal data concerning him or her without undue delay and the controller shall have the obligation to erase personal data without undue delay..."

In simpler terms, this regulation empowers individuals with the right to request the removal of their personal data from an organisation's records. Upon receiving such a request, the organisation is legally obligated to delete the specified personal data promptly, without unnecessary delay [9]. Data deletion on request also applies to CCPA [17].

5.4.1 Encryption

Encryption and crypto-shredding are viable techniques to support data localization laws, improve data protection, and ensure compliance with regulations such as the one above from GDPR [39].

Although encryption is the process of converting plaintext information into an unreadable format (ciphertext) using a cryptographic algorithm and a key [39] [40]. Crypto-shredding is the process of destroying data by destroying the cryptographic keys that protect the data [42]. This renders the encrypted data useless, effectively erasing them

without physically deleting every instance in the organisation's database. Only those with the correct decryption key can access the original information again [39]. Encrypting data is a very effective way to protect it during transfer and a reliable method of securing stored personal data. It also mitigates the risk of internal misuse, as only authorized individuals with the correct key can access the data [39].

In encryption schemes, two techniques can be employed to ensure data security: Symmetric encryption and asymmetric encryption [40]. The main difference between these two types of encryption is that symmetric encryption uses the same key for both encryption and decryption [40].

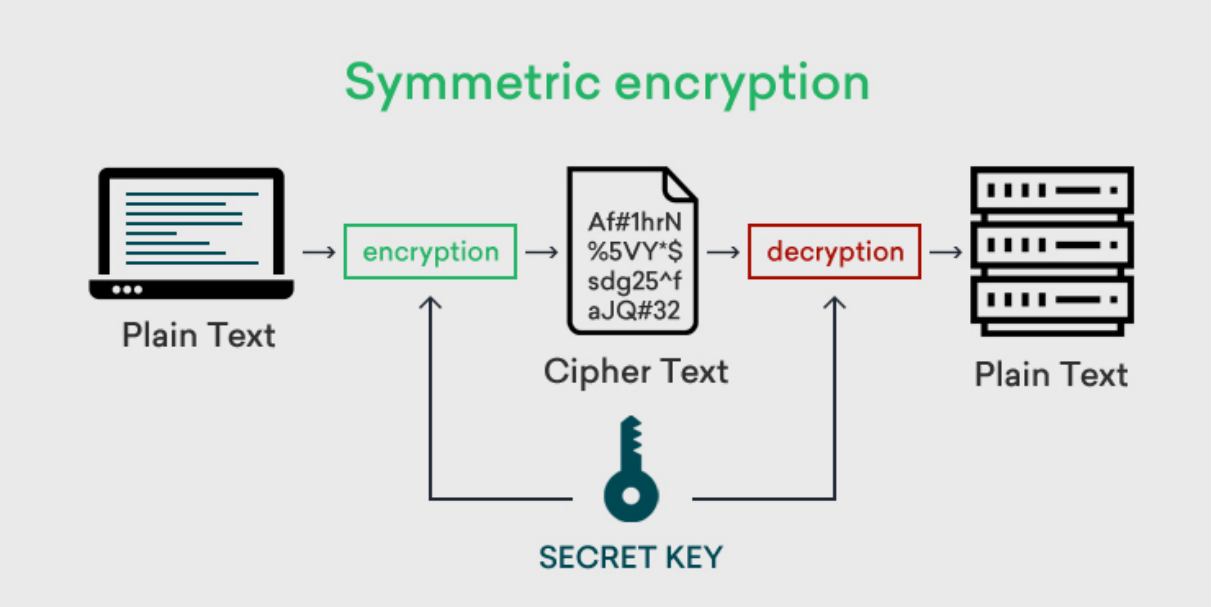


Figure 5.1: Symmetric encryption; Source: ClickSSL [40]

In contrast, asymmetric encryption utilizes two distinct keys: a public key for encrypting the data and a private key for decrypting it [40].

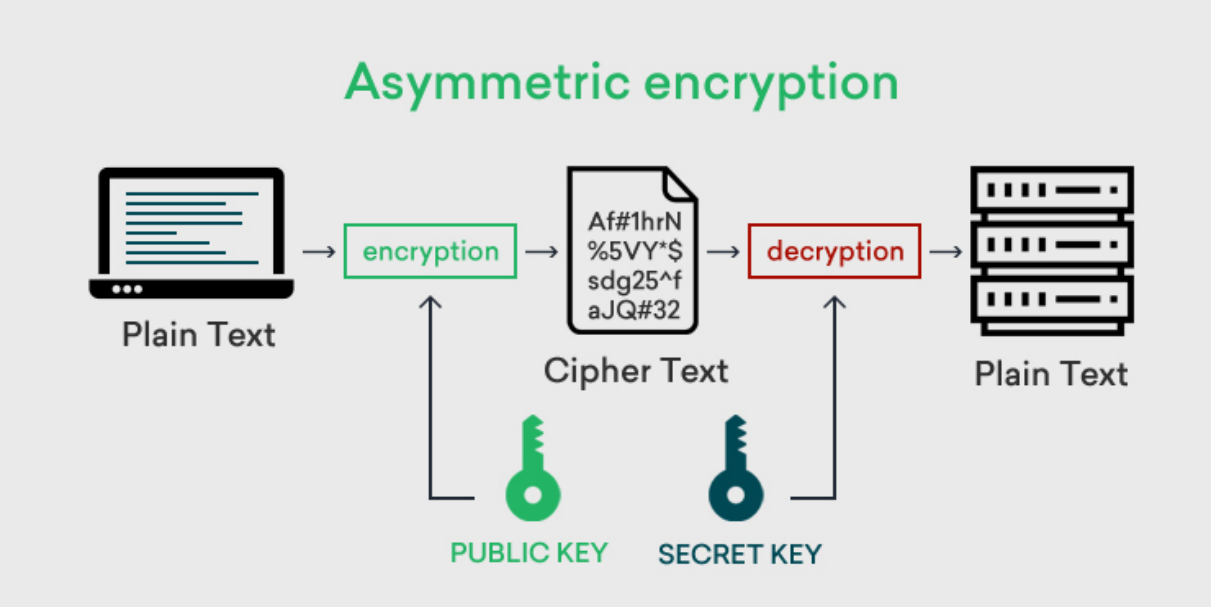


Figure 5.2: Asymmetric encryption; Source: ClickSSL [40]

Neither the GDPR nor the CCPA prescribe specific encryption techniques, but they both recommend using appropriate technical and organisational measures to ensure data security, which includes encryption. For example, GDPR Article 6 identifies "encryption

or pseudonymisation" as "appropriate safeguards" for protecting subjects' personal data [17]. GDPR Article 32 states that "the processor shall implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk, including: (a) the pseudonymisation and encryption of personal data..." [17]. And Article 34, in communication of a personal data breach to the data subject, allows organisations suffering a data breach to avoid the communication requirement if they used encryption to "render the personal data unintelligible to any person unauthorised to access it" [17]. Section 1798.150 of the CCPA allows consumers to take legal action against a business if their nonencrypted and nonredacted personal information is accessed, stolen, or disclosed without authorization due to the business's failure to implement and maintain reasonable security measures [17].

Although both encryption methods can be used, symmetric encryption seems to be simpler as only one key is needed. For encrypting bulk data and data at rest, symmetric cryptography is preferred because of it's high speed and swifter execution functionalities [40] [41]. Popular symmetric encryption algorithms are: AES, QUAD, RC4, 3DES and DES [40]. To encrypt communications between systems for data in transit, SSL (Secure Sockets Layer) and Transport Layer Security (TLS) protocols should be employed [41]. SSL/TLS uses both asymmetric and symmetric encryption techniques to safeguard the confidentiality and integrity of data in transit [43]. Popular asymmetric encryption algorithms are: DSA, RSA, EL GAMAL, ECC and Diffie Hellman [40].

In addition to traditional encryption, hashing and end-to-end encryption (E2EE) are other essential techniques for securing data that can be used to protect user's data. Hashing converts data into a fixed-size string, creating a unique fingerprint. It's commonly used for securing passwords because it's challenging to reverse-engineer the original data. For example, SHA-256 is a widely used hash function [41]. Then we have E2EE that encrypts data by the sender and only decrypts it by the recipient, preventing intermediaries from accessing the data during transmission. This method is vital for messaging apps to maintain private communications [41]. E2EE often combines both asymmetric and symmetric encryption techniques, but primarily asymmetric encryption.

5.5 Evaluations and Discussion

As we have seen, data localization laws have a significant impact on their environment, particularly concerning consumers, businesses and global trade and economics. The GDPR's reach is extensive as it applies not only to companies based in the EU but also to all companies that process EU data, regardless of their location. Although the CCPA applies only to one US state, California, it will impact over 500,000 businesses [23], in addition to the global influence of GDPR on companies.

Ensuring compliance is often an ongoing and costly task. However, investing in data localization laws compliance usually pays off because the costs are lower than potential fines. The GDPR imposes higher fines overall compared to the CCPA [15] [17].

Although there is a distinction between personal data and personal information, CCPA safeguards personal information of California residents, as well as their households and devices, while GDPR protects natural persons, EU citizens and residents, both want to protect their parties from unlawful data collecting, processing, trading, and sharing through stringent regulations. Both want to strengthen the rights of their citizens and residents, and want to help with data protection.

Many organisations find it challenging to become compliant with these regulations, often due to a lack of understanding of what is required [32]. Compliance can be achieved through various means or combinations of them, such as obtaining GDPR certificates [20], using encryption [41], leveraging the latest technologies [21], and providing staff training

[21]. Studies show that 90 % of data breaches involve human error [21], underscoring the importance of a comprehensive compliance strategy that includes both technological solutions and process improvements.

The paper "The 10 Problems of the GDPR" by Roslyn Layton from March 12, 2019, argues that these regulations are not only for protecting it's citizens and residents but also for increasing the governmental power under the guise of customer control [29]. This is a provocative statement that I would like to leave as is. It can be used to stimulate further discussions and debates.

5.6 Summary and Conclusions

Data protection laws have profound implications for both consumers and businesses. They enhance security and give consumers more control over their personal data, but also introduce complexities and increase costs. These laws can affect global service availability, leading to inconsistent experiences and economic losses due to Internet fragmentation.

For businesses, compliance can reduce data breaches and improve reputation, but involves significant financial and operational challenges. High-profile fines highlight the importance of compliance, which requires substantial investments in technology, staff training, and data management.

Data localization laws aim to protect personal information and data, but can hinder international trade by restricting data flows, leading to inefficiencies, increased costs, and reduced innovation. These laws can decrease market competition, especially impacting small and medium-sized enterprises and leading to market exits due to high compliance costs. They also pose challenges for fields reliant on large data sets and advanced technologies.

In conclusion, while data protection and data localization laws aim to enhance data protection, they introduce significant challenges that can impact global trade, economic growth, and technological innovation. Balancing personal data protection with the needs of a dynamic global economy is essential. Techniques like encryption, crypto-shredding, hashing, and end-to-end encryption are crucial and help protect personal data and ensure compliance with these laws.

Bibliography

[1] *Volumen der jährlich generierten/replizierten digitalen Datenmenge weltweit von 2010 bis 2022 und Prognose bis 2027*; https://de.statista.com/statistik/daten/studie/267974/umfrage/prognose-zum-weltweit-generierten-datenvolumen/, October 2024.

[2] *Marriott zahlt nach Data Breaches 52 Millionen Dollar Entschädigung*; https://www.inside-it.ch/marriott-zahlt-nach-data-breaches-52-millionen-dollar-entschaedigung-20241011, October 2024.

[3] *New Federal Act on Data Protection (nFADP)*; https://www.kmu.admin.ch/kmu/en/home/facts-and-trends/digitization/data-protection/new-federal-act-on-data-protection-nfadp.html, October 2024.

[4] *Art. 2 GDPR; Material scope*; https://gdpr-info.eu/art-2-gdpr/, October 2024.

[5] *Art. 4 GDPR; Definitions*; https://gdpr-info.eu/art-4-gdpr/, October 2024.

[6] *Chapter 3; Rights of the data subject*; https://gdpr-info.eu/chapter-3/, October 2024.

[7] *What is GDPR, the EU's new data protection law?*; https://gdpr.eu/what-is-gdpr/, November 2024.

[8] *Art. 9 GDPR; Processing of special categories of personal data*; Nr. 1, https://gdpr-info.eu/art-9-gdpr/, October 2024.

[9] *Right to erasure ('right to be forgotten')*; https://gdpr.eu/article-17-right-to-be-forgotten/, October 2024.

[10] Ben Wolford: *Does the GDPR apply to companies outside of the EU?*; https://gdpr.eu/companies-outside-of-europe/, October 2024.

[11] Ben Wolford: *Everything you need to know about the GPDR Data Protection Officer (DPO)*; https://gdpr.eu/data-protection-officer/, October 2024.

[12] *What is GDPR?*; https://doc.milestonesys.com/2020R1/en-US/quick_guides/gdprprivacyguide/gdpr_whatisgdpr.htm, October 2024.

[13] *California Consumer Privacy Act (CCPA)*; https://oag.ca.gov/privacy/ccpa/, March 2024.

[14] *The California Privacy Rights Act of 2020*; https://thecpra.org/, 2020.

[15] *CCPA vs GDPR*; https://www.cookiebot.com/en/ccpa-vs-gdpr/, May 2024.

[16] *CCPA vs GDPR. What's the Difference?*; https://www.cookieyes.com/blog/ccpa-vs-gdpr/, November 2024.

[17] *CCPA vs GDPR Compliance Comparison*; https://www.entrust.com/resources/learn/ccpa-vs-gdpr, November 2024.

[18] *The CCPA vs. the GDPR comparison*; https://www.onetrust.com/blog/the-ccpa-vs-the-gdpr/, December 2019.

[19] *20 biggest GDPR fines so far [2024]*; https://dataprivacymanager.net/5-biggest-gdpr-fines-so-far-2020/, November 2024.

[20] *Compliance Q&A: How much does GDPR compliance cost?*; https://sprinto.com/blog/gdpr-compliance-cost/, November 2024.

[21] *How Much Does GDPR Compliance Cost in 2023?*; https://www.itgovernance.eu/blog/en/how-much-does-gdpr-compliance-cost-in-2020, May 2023.

[22] *Impact of Data Localization Requirements on Commerce and Innovation*; https://www.americanactionforum.org/insight/impact-of-data-localization-requirements-on-commerce-and-innovation/, November 2024.

[23] Louise: *The differences between GDPR and CCPA and what they mean for ecommerce*; https://blog.carts.guru/ccpa-gdpr-ecommerce, January 2020.

[24] *How much does ISO 27001 certification cost?*; https://www.vanta.com/collection/iso-27001/iso-27001-certification-cost, November 2024.

[25] Christopher Bret Alexander: *The General Data Protection Regulation and California Consumer Privacy Act: The Economic Impact and Future of Data Privacy Regulations*; Loyola Consumer Law Review, Volume 32, Issue 2, 2020, https://lawecommons.luc.edu/lclr/.

[26] Raffaele Congiu, Lorien Sabatino, Geza Sapi: *The Impact of Privacy Regulation on Web Traffic: Evidence From the GDPR.*; Information Economics and Policy, December 2022, https://www.sciencedirect.com/science/article/abs/pii/S0167624522000427.

[27] Seo, Junwoo, Kyoungmin Kim, Mookyu Park, and Kyungho Lee: *An Analysis of Economic Impact on IoT Industry under GDPR*; Mobile Information Systems, 2018, https://doi.org/10.1155/2018/6792028.

[28] Roslyn Layton, Silvia Elaluf-Calderwood: *A Social Economic Analysis of the Impact of GDPR on Security and Privacy Practices*; Center for Communication, Media and Information Technologies, Aalborg University and Delray Beach, United States of America, 2019, https://www.privacysecurityacademy.com/wp-content/uploads/2019/08/A-Social-Economic-Analysis-of-the-Impact-of-GDPR-on-Security-and-Privacy-Practices.pdf.

[29] Roslyn Layton: *The 10 Problems of the GDPR, The US can learn from the EU's mistakes and leapfrog its policy*; Statement before the Senate Judiciary Committee, On the General Data Protection Regulation and California Consumer Privacy Act: Opt-ins, Consumer Control, and the Impact on Competition and Innovation, March 12, 2019, https://www.judiciary.senate.gov/imo/media/doc/Layton%20Testimony1.pdf.

[30] *Commission proposes a comprehensive reform of data protection rules to increase users' control of their data and to cut costs for businesses*; https://ec.europa.eu/commission/presscorner/detail/en/ip_12_46, October 2024.

[31] *General conditions for imposing administrative fines*; Parliament and Council Regulation 2016/679, Art. 83, Nr. 5., L 119/83, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679.

[32] *Organizations Worldwide Fear GDPR Non-Compliance Could Put Them Out of Business*; VERITAS Technologies LLC, https://uk.insight.com/content/dam/insight/EMEA/blog/2017/06/GDPR-Infographic-design-final.pdf, November 2024.

[33] Matt Buckley, *Federal Data Privacy Regulation: Do Not Expect an American GDPR*; 21 DePaul Bus. & Com. L.J. (2023), pp. 156, https://via.library.depaul.edu/bclj/vol21/iss2/6.

[34] *General conditions for imposing administrative fines*; Parliament and Council Regulation 2016/679, Section 4, Art. 37-39, L 119/55-56, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679.

[35] I. van Ooijen & Helena U. Vrabec: *Does the GDPR Enhance Consumers' Control over Personal Data? An Analysis from a Behavioural Perspective*; Journal of Consumer Policy 42 (6221), https://www.researchgate.net/publication/329570309_Does_the_GDPR_Enhance_Consumers'_Control_over_Personal_Data_An_Analysis

[36] Garrett Johnson: *ECONOMIC RESEARCH ON PRIVACY REGULATION: LESSONS FROM THE GDPR AND BEYOND*; NBER WORKING PAPER SERIES, Working Paper 30705, December 2022, pp. 5-9, http://www.nber.org/papers/w30705.

[37] *DATA PROTECTION & GDPR*; Prisoners' Advice Service - Information Sheet, pp. 1-2, https://www.prisonersadvice.org.uk/wp-content/uploads/2021/05/Data-Protection-2018-with-added-GDPR.pdf.

[38] *Temporary Specification for gTLD Registration Data*; ICANN, https://www.icann.org/resources/pages/gtldregistration-data-specs-en, May 17, 2018

[39] *GDPR; Encryption*; https://gdpr-info.eu/issues/encryption/, November 2024.

[40] *Symmetric vs Asymmetric Encryption - What Are the Difference?*; https://www.clickssl.net/blog/symmetric-encryption-vs-asymmetric-encryption, November 2024.

[41] *A Guide to GDPR Data Encryption*; https://www.gdpr-advisor.com/a-guide-to-gdpr-data-encryption/, November 2024.

[42] Brent Robinson: *Crypto shredding: How it can solve modern data retention challenges*; https://medium.com/@brentrobinson5/crypto-shredding-how-it-can-solve-modern-data-retention-challenges-da874b01745b, January 2019.

[43] *What is SSL/TLS Encryption?*; https://www.f5.com/glossary/ssl-tls-encryption, November 2024.

Chapter 6

An Overview of Sustainable AI Regulations

Panagiotopoulou Maria Christina & Urech Rafael

This report examines current regulations aiming at the sustainability of artificial intelligence (AI), with a primary focus on its role in addressing environmental, social, and ethical challenges. It discusses ways to reduce AI's environmental impact and harness its potential to address global issues like climate change and inequality. The report reviews methods for evaluating sustainability, such as life cycle assessments and ethical metrics, and analyzes regulatory frameworks like the EU AI Act. It highlights both the benefits and risks of AI and offers recommendations for green AI practices, ethical standards, and global collaboration to promote fair and responsible AI development.

Contents

6.1 Introduction

6.1.1 Context & Importance of Sustainable AI

The rapid advancement of artificial intelligence (AI) has revolutionized numerous aspects of society, offering unprecedented opportunities for innovation while introducing complex challenges. Among these, sustainability has emerged as a critical concern, encompassing environmental, social, and ethical dimensions. The substantial energy consumption of AI systems, their potential to exacerbate existing inequalities, and the ethical dilemmas they pose underscore the need for sustainable development and governance of AI. Achieving sustainability in AI requires balancing technological progress with the responsibility to mitigate environmental impacts, promote equitable access, and uphold societal values.

6.1.2 Purpose of the Report

The purpose of this report is to analyze sustainable AI regulations, looking at their current state, the challenges they face, and possible future developments. It explores how regulatory frameworks, ethical principles, and societal needs come together to balance the benefits of AI with its potential risks. By reviewing existing regulations, evaluating how well they work, and identifying any gaps, the report aims to offer insights into building better governance systems for sustainable AI.

6.1.3 Scope and Structure of the Report

This report examines sustainable AI through its environmental, social, and ethical dimensions, emphasizing the need to balance technological advancement with responsibility. It begins by conceptualizing sustainability in AI, distinguishing between reducing AI's impact and leveraging it for broader societal benefits. Metrics and methods such as Life Cycle Assessments, energy consumption, and carbon emissions are explored to assess environmental sustainability.

The analysis then addresses AI's societal impacts, focusing on biases in algorithms, opaque decision-making, and economic inequities. These issues are examined alongside their root causes and implications for individuals and communities. Regulatory frameworks, including the General Data Protection Regulation (GDPR), the EU AI Act, and the United Nations Sustainable Development Goals (SDGs), are reviewed to highlight their strengths and areas for improvement.

The report concludes with recommendations for ethical and sustainable AI practices, such as enhancing transparency, ensuring accountability, and fostering global collaboration. These discussions aim to provide actionable insights for aligning AI development with environmental, ethical, and social priorities, offering a comprehensive framework for advancing sustainable AI.

6.2 Sustainability in AI

6.2.1 Conceptualizing Sustainability in Artificial Intelligence

Sustainability in artificial intelligence refers to the responsible development and deployment of AI systems that minimize negative impacts while maximizing societal and environmental benefits. This concept captures a dual focus: reducing the resource-intensive nature of AI technologies and leveraging AI to address pressing global challenges such as climate change, inequality, and ethical dilemmas.

At its core, sustainability in AI recognizes the interconnectedness of technological innovation with broader ecological and social systems. It emphasizes the need to evaluate AI's impact across its entire lifecycle—from the extraction of raw materials for hardware to the energy demands of training large models and the societal consequences of their deployment. Sustainable AI development thus entails designing systems that are energy-efficient, ethically grounded, and inclusive.

This framework extends beyond environmental considerations to address social equity and governance. Ensuring equitable access to AI technologies and preventing their misuse or unintended consequences are key aspects of sustainability. For instance, AI systems must be designed to respect human rights, avoid perpetuating biases, and align with ethical standards. Equally, they should enhance opportunities for underserved communities, bridging rather than widening digital divides.

Conceptualizing sustainability in AI requires a forward-looking perspective that anticipates long-term impacts while addressing immediate concerns. By adopting this approach, AI developers, policymakers, and stakeholders can ensure that AI technologies contribute to a more equitable and sustainable future, rather than exacerbating existing challenges. This broad yet focused understanding sets the stage for exploring specific dimensions of sustainability in AI, including its environmental, social, and ethical implications.

6.2.2 Distinguishing Sustainable AI and AI for Sustainability

Sustainable AI and AI for sustainability represent complementary approaches to technological development. Van Wynsberghe (2021) provides a clear distinction between these concepts, emphasizing that Sustainable AI focuses on reducing the negative environmental and social impacts of AI technologies, such as energy consumption and ethical dilemmas. On the other hand, AI for Sustainability highlights the potential of AI to address systemic global challenges, including climate change, resource management, and social equity [3].

This distinction underscores the dual role of AI in achieving a balance between mitigating its own footprint and enabling solutions to complex problems. By integrating these perspectives, policymakers and developers can align AI development with broader sustainability goals.

6.2.3 Integrated Environmental, Computational, Social, and Ethical Dimensions

Artificial intelligence (AI) holds immense potential to transform society, but its sustainability depends on addressing its environmental, computational, social, and ethical implications. A critical concern is the environmental footprint of AI systems, particularly the substantial computational resources required for training large models. Freitag et al. emphasize the climate impact of information and communication technologies, demonstrating the need to consider the energy intensity and associated carbon emissions of advanced AI systems [14].

Beyond environmental concerns, AI's computational dimension offers opportunities to address pressing global challenges. Lacoste et al. highlight how machine learning can be strategically deployed to combat climate change by optimizing energy use and reducing inefficiencies in various sectors. This approach underscores the importance of designing computational methods that not only solve problems but also prioritize ecological responsibility [18].

Social and ethical dimensions are equally integral to AI sustainability. These dimensions focus on ensuring that AI systems are developed and deployed to promote human well-being, uphold individual rights, and prevent societal harm. Achieving this requires

a holistic perspective that addresses potential biases, protects vulnerable groups, and ensures equitable access to AI technologies.

By integrating these dimensions, sustainability in AI becomes a multidisciplinary endeavor. Collaboration between technologists, environmental scientists, ethicists, and policymakers is essential to create AI systems that are not only technologically advanced but also environmentally responsible and socially inclusive. Such an approach ensures that AI contributes positively to global challenges without exacerbating existing inequalities or environmental issues.

6.2.4 Towards a Holistic Approach

Sustainability in artificial intelligence represents a critical challenge for technological development. It demands a comprehensive approach that balances innovation with environmental responsibility, social equity, and ethical considerations. Bommasani et al. highlight that sustainability in AI must transcend traditional technological development, integrating technological capabilities with environmental and social dimensions to address both the transformative potential and inherent risks of advanced computational systems [16].

This holistic perspective requires engaging multiple disciplines—computer science, environmental research, social sciences, and policymaking—to ensure that AI systems contribute positively to global challenges. Such integration fosters technological advancements that are not only efficient but also aligned with societal values and ecological priorities. Continuous research, collaborative efforts, and proactive governance will be essential to realizing AI's potential while mitigating its negative impacts.

6.3 Methods and Metrics to asses Sustainability

To measure the sustainability of AI, one needs units of measurement to enable comparisons between the "before" and "after" states. However, due to the diverse and multifaceted areas impacted and influenced by AI, measuring sustainability presents a high degree of complexity. Before working on measuring the sustainability of AI, we would like to point out the challenges found in attempting to measure the sustainability of computing, as well as the difficulty of finding metrics that make social and ethical development measurable. Given the focus of this report on environmental, social, and ethical sustainability, metrics related to these areas will be covered in this chapter.

6.3.1 Measuring Environmental Sustainability

To measure the environmental sustainability of any technology, the environmental impact of said technology must be assessed. This impact is multifaceted. The environment can be affected in various areas, such as climate, terrestrial, or aquatic ecosystems. Metrics for measuring such impacts include carbon emissions, air and water pollution, energy and water consumption, and biodiversity affected by technology [34, p.6-11].

A commonly used method to quantify environmental impacts from technologies is the Life Cycle Assessment (LCA). The LCA assesses the impact of a product or service throughout its entire life cycle. Guidelines for conducting LCAs have been developed and issued by the International Organization for Standardization (ISO 14040).

The assessment consists of four interconnected phases, beginning with defining the goal and scope of the study, which establishes the system boundaries, study duration, and a

functional unit for comparison. The second and main step is the life cycle inventory (LCI), where all inputs and outputs within the defined boundaries are quantified. Inputs include raw materials and energy, while outputs consist of products, waste, and emissions to air, water, and soil. The accuracy of the LCA heavily depends on this phase. The results from the LCI are then classified into environmental impact categories, such as global warming, acidification, eutrophication, and ozone layer depletion. The final step is the continuous interpretation and evaluation of results [37].

One could assume that following this standardized approach should render clear results regarding the environmental impact of artificial intelligence and make a meaningful conclusion about its sustainability possible. However, assessing the environmental impact of any technology, especially rapidly evolving, cutting-edge technology, proves difficult. This challenge is identical to those faced in addressing sustainability in computer science. The use of an LCA is neither new nor limited to the field of artificial intelligence. An article published in 2004 by Andreas Köhler and Lorenz Erdmann addressed the expected environmental impact of pervasive computing using a combination of different methods, including the Life Cycle Assessment (LCA), to address the uncertainties inherent in emerging technologies. Köhler and Lorenz were face with multiple challenges trying to assess the enviromental impact of pervasive computing identified four main difficulties analyzing cutting edge technology with the Life Cycle Assessment.[36]

1. **Data Uncertainty.** Analysis regarding microelectronics face significant data challenges due to the complexity of the production processes. In their example the dynamic and global Supply Chain involved over 400 steps and dynamic global supply chains. Existing studies provide only simplified LCAs, which are insufficient for comprehensive assessments [36, p.833-834].

2. **Usage Uncertainty.** Pèredicting the environmental impact of pervasive computing is hindered by uncertainties about how technologies will develop and how they will be used in the future. The lack of knowledge about future usage patterns makes modelling environmental impacts accurately difficult, especially for emerging technologies. applications [36, p.834].

3. **Inadequate System Boundaries.** LCA methodologies have difficulties capturing the broad and dynamic nature of pervasive computing technologies. Narrowly defined boundaries can omit significant impacts and fail to reflect the interconnected nature of these system. [36, p.833-834].

4. **Rebound effects.** The potential environmental benefits of in Köhlers adn Erdmann example pervasive computing, such as improved efficiency and dematerialization, are often offset by rebound effects, where increased demand and new use cases neutralize anticipated savings, by increasing demand. [36, p.832-834].

These issues illustrate the struggles of completely capturing the environmental impact of any technology. Not being able to know future outcomes, make it impossible to know the effective impact in advance, which is especially true for evolving, dynamic and cutting edge technologies. Unknow factors such as the rebound effect can decrease enviromental feasabilty, however future developments can also impro enviromental sustainability.

Köhler and Erdmann mention that Embedding electronics into non-ICT objects complicate recycling processes, causing cross-contamination in waste streams and increasing the loss of valuable materials. Current recycling systems are according to them often

unsuitable for managing these challenges, leading to potential environmental harm[36, p.833-834]. However since the publishing of the paper recycling Technology has advanced significantly, strongly influence by embedding ICT Objects and their advantages in the Recycling [17], thereby reducing the enviromental impact of Perpervasive computing.

Why was a Paper from 20 years ago chosen, to elaborate on the challenges of capturing the entire enviromental impact of a currentlys evoloving technology, in our case Artificial Intelligence. Because Köhler and Erdmann were stanmding in the same spot regarding pervasive computing as we are currently with Artificial Intelligence. The difficulties they had in capturing the entire impact because of the dynamic and evolving nature of cutting Edge Technology can be seen in trying to measure the enviromental Impact of Aritificial Intellingence aswell.

Current research is struggeling with similar issues, as Köhler and Erdmann did 20 years ago. Ligozat et al. are faced with multiple challenges assessing the enviromental Impact of AI Solutions.[37] Main Challenges among other Things Scope of Definition, Incomplete Life Cycle Coverage, Data Availability, indirect and third Order effects, Complexity of AI Systems, uncertain environmental gains and Dynamic Nature of AI Technologies. Additionally the balancing of multiple criteria Enviromental impacts span diverse categories, such as carbon footprint, resource depletion, and human toxicity. Balancing and aggregating these criteria into a coherent evaluation framework is difficult [37]. To elaborate further on the challenges of being able to capture the environmental impact We will look at certain findings of Ligozat et al.

Köhler and Erdamnn note that an inventory analysis of microchip production had to deal with more than 400 processes [36, p.833]. Ligolzat et al a noted lack of Life Cycle Assessment studies for the production phase of GPUs or TPUs and reliable Data thereof. A cited study for a CPU-only data center in France revealed that 40 % of its greenhouse gas emissions stemmed from the production phase, underlining the importance of inclusion of such costs in a assesment. [37, p.8-8].]. Graphical Processing Units (GPUs) and Tensor Processing Units (TPUs) are detrimental to the field of deep learning and, therefore, artificial intelligence, due to their vastly better performance compared to CPUs attributed to their domain-specific hardware design [38], makei a lack of reliable data even more critical. Galindo Serrano et al. [39] aswell as Guldbrandsson and Bergmark [40] underline the importance of minimizing uncertainty in the data used in Life Cycle Assessment. It is crucial to ensure validity. This however proves difficult as elaborated by Ligozat et al.

The Complexity of AI systems aswell as the allocation of shared resources complicate the assesment. Servers in data centres are used for multiple purposes simultaneously. Ligozat et al. discusses methods to allocate environmental costs proportionally, by execution time of the AI service [37, p.6-8]. This would however again encompass costs from production, over use up to the end of life.

Further difficulties arise from indirect and third-Order Effects. The rebound effect, already discussed by Köhler and Erdmann, where increased efficiency leads to more consumption, is diffictult to predict. Liozat et al use a smart building where energy-saving AI might lead users to increasing their thermostat settings due to perceived savings and thereby increasing energy use overall, as example[37, p.8-9]. Yet uncertain enviromental gains could however reduce the cost. An analysis recent research regarding AI for Green applications revealed that while AI certainly could optimize energy use the full environmental impact, including hardware production and operation, often remains unquantified in current research[37, p.8-10].

Working out all these difficulties one is faced trying to measure the enviromental impact, one can conclude, that currently, Complexity of Systems, lack of Data aswell as unpredicatble future influences make assesing the sustainability of AI very difficult since capturing the whole impact seems impossible and finding metrics to measure said sustainability strongly depends on which phase is to be assesed and reduction of metrics poses the risk of ovrsimplifying the model.

This leads to an observable problem in current research regarding the environmental impact of AI. The lack of data, forces researchers to assess the environmental impact of AI using the data available, which, if available at all, consists of the energy and water consumption from the use and inference phases. The carbon footprint of the use and inference phase can be calculated by using the carbon emission output of electricity production depending on the region.

6.3.2 Measuring Social Sustainability

Tackling the task of measuring social sustainability is far from trivial. Firstly, social sustainability needs to be defined. Social sustainability and ethical sustainability are strongly intertwined. Terminology such as sustainability, responsibility, and ethics is often used synonymously and therefore incorrectly, leading to confusion [24].

To be able to measure social sustainability, it is necessary to define it and distinguish it from ethical sustainability. The concepts of social and ethical sustainability, while interconnected, address distinct facets of sustainable development. Social sustainability focuses on the structures and processes that support the well-being of individuals and communities, emphasizing equity, diversity, quality of life, social capital, and community development [25]. In contrast, ethical sustainability pertains to the moral principles guiding human interactions with each other and the environment, encompassing issues like justice, rights, duties, and moral obligations [26]. The broadest way to define social sustainability is as the impact it has on people [27]. Therefore, measuring social sustainability requires assessing the positive or negative impact artificial intelligence (AI) has on people's lives. Adding to the complexity is the challenge of doing so using comparable metrics.

Measuring the social sustainability of AI and other technologies presents significant challenges due to the complex and multifaceted nature of social impacts. While current research touches on the social sustainability of AI, the primary focus still lies on the environmental aspects of AI and sustainability.

Thelisson et al. created a dashboard encompassing multiple dimensions of sustainability for AI and proposed various indicators for evaluating the social sustainability of AI, including well-being monitoring of employees, impact assessments of AI systems, risk assessments, and other metrics [28]. However, these proposed indicators represent high-level concepts rather than concrete, applicable measurements. From these proposals, effective metrics are still a distant goal.

Similar challenges are observable in the work of Kumar et al. Their study provides insights into the trade-offs and challenges in assessing the sustainability of AI-based systems. Kumar et al. conclude that there is insufficient holistic coverage of potential sustainability benefits or costs [29].

This showcases a clear understanding of the social impact of AI and the necessity of

making this impact quantifiable to assess the social sustainability of AI. However, quantifying such impact poses challenges, as most proposed indicators and metrics remain high-level categories and lack concrete measurements.

An already existing framework for measuring social well-being is the OECD's How's Life? report, which is based on a continuously evolving dataset. This report analyzes global well-being using over 80 indicators, ranging from homicide rates, life expectancy at birth, and PISA scores to housing affordability, the gender wage gap, and household income. These indicators create a framework for assessing well-being and interpreting data using an extensive dataset that provides comparability over decades [30].

By integrating multiple dimensions, the How's Life? report provides a holistic perspective on social sustainability [30]. However, applying this extensive framework to measure the impact of AI on any of these indicators is impractical due to the multifaceted nature of the indicators themselves. Narrowing down any change to a single contributing factor, such as AI, is impossible, making a generalized measurement of the social sustainability of AI nearly unattainable. This is due to the absence of an assessment framework capable of addressing the broad scope of potential metrics.

In conclusion, measuring the social sustainability and social impact of AI is currently limited to case studies, where individual applications of AI are analyzed for their impact in specific dimensions, such as job creation or education. A generalized framework with clear metrics for assessing social sustainability remains lacking.

6.3.3 Measuring Ethical Sustainability

The previously mentioned issue that social and ethical sustainability are strongly intertwined and often used synonymously insinuates an overlap regarding their measurability and general metrics. However, as previously covered, the social aspect in this context is defined by the impact AI has on people's lives. Ethical sustainability, meanwhile, also involves measuring the ethical behavior of AI.

Van Wynsberghe splits the concept of "ethical sustainability" regarding artificial intelligence into two parts. The first part concerns the ethical use of AI, meaning how end users decide to use and apply artificial intelligence. The second part focuses on ensuring that AI systems themselves are developed and deployed in ways that uphold ethical standards [3]. We agree with this separation, and for the purpose of measurability and metrics, we will focus on the ethical sustainability of the models themselves.

When attempting to assess the ethical sustainability of AI models regarding their behavior, a separation into higher-level principles (to define what needs to be addressed) and measurable, concrete metrics appears to be a reasonable approach. The High-Level Expert Group on AI of the European Commission outlined seven ethical principles in their 2019 Ethics Guidelines for Trustworthy AI [31]:

1. **Human Agency and Oversight**
2. **Technical Robustness and Safety**
3. **Privacy and Data Governance**
4. **Transparency**
5. **Diversity, Non-Discrimination, and Fairnes**

6. **Societal and Environmental Well-Being**

7. **Accountability**

These principles are non-technical, broadly understandable, and cover various aspects of ethical AI. Palumbo et al. conducted a systematic literature review to identify and categorize current metrics for ethical AI development. Examining 66 articles from 2018 to 2023, they analyzed the frequency with which each ethical principle was addressed in the literature. The ethical principles outlined in the Ethics Guidelines for Trustworthy AI served as categories. Their review highlighted that the most addressed principles were Diversity, Non-Discrimination, and Fairness (58%), followed by Transparency (37%), while other principles were underrepresented. Furthermore, of the articles reviewed, only 12 provided actual objective and measurable metrics. These measurable metrics related solely to the principle of "Diversity, Non-Discrimination, and Fairness," limiting the scope of ethical metrics for measuring ethical sustainability to a single principle [32].

Although Palumbo et al. emphasize the need for practical, objective tools to monitor and enhance the ethical compliance of AI systems throughout their lifecycle, their research reveals that, in contrast to social sustainability, there are metrics available for measuring the ethical sustainability of AI models. However, a general analysis of any AI model using the same metrics is, in our opinion, impractical. Different metrics must be individually weighted according to the AI use case. For instance, the false positive rate (FPR) for AI-supported loan approval is less critical than the false positive rate for AI-predicted reoffending rates of criminals. A wrongly approved loan is less consequential than being denied parole due to a false positive, thereby extending an individual's imprisonment.

For a detailed list of the gathered metrics, we refer to the literature review itself. However, a brief, non-exhaustive overview of some metrics is provided. The false positive rate (FPR) and false negative rate (FNR) are used to evaluate fairness by assessing discrepancies between groups [32, p.6-7]. FPR represents the proportion of incorrect positive predictions among all actual negatives, while FNR represents the proportion of incorrect negative predictions among all actual positives. Importantly, these metrics do not represent overall accuracy but must be assessed regarding discrepancies between different groups. A model may have high overall accuracy, implying technical robustness, yet variations in FPR and FNR between privileged and unprivileged groups raise red flags regarding ethical sustainability. Fairness is achieved if the FPR and FNR values are equal across these groups [32, p.6-7, p.16-17].

Pokholkova et al. propose the concept of expert workshops for quantifying adherence to ethical principles. Metrics include weighted sums of expert judgments on criteria such as documentation of decision-making processes, traceability of data through AI models, and user comprehension of the system. These criteria are assessed through both qualitative and quantitative phases of workshops [33, p.5-7]. However, reliance on expert opinions introduces the risk of bias, as present biases among experts may be projected onto the decision-making process [33, p.6].

This review of current challenges in using metrics to assess the ethical sustainability of artificial intelligence demonstrates that, unlike the difficulty of measuring social sustainability, there are metrics available for evaluating the ethical behavior of AI models. However, the current focus is heavily skewed toward Diversity, Non-Discrimination, and Fairness, with other principles remaining underrepresented.

6.4 Social Impact of AI in Society

6.4.1 Dual Impact of AI in Society

The integration of Artificial Intelligence (AI) into modern society presents a complex duality of transformative benefits and significant challenges. Understanding this duality requires careful analysis of the opportunities and risks that AI technologies present across various sectors.

On the negative side, AI has profoundly altered social dynamics. Increased reliance on AI-mediated communication and virtual interactions risks diminishing face-to-face social skills and creating echo chambers that limit exposure to diverse perspectives. These effects are particularly evident in social media algorithms and content recommendation systems that can reinforce polarization and reduce social cohesion [22].

The advancement of AI also raises concerns about job displacement. While automation initially targets routine tasks, even knowledge-based professions are increasingly affected by AI systems capable of complex decision-making. Workers in sectors without robust opportunities are especially vulnerable, potentially exacerbating socioeconomic disparities [12]. Additionally, the unequal distribution of AI's benefits risks widening existing wealth gaps, as technological and financial resources become concentrated in AI-capable entities and regions, creating new forms of economic disparity.

Bias and discrimination are critical concerns as well. AI systems trained on historical data often perpetuate societal prejudices, resulting in discriminatory outcomes in areas such as hiring and lending. This highlights the need for transparent and equitable AI systems to prevent the institutionalization of bias at scale [3].

Despite these challenges, the positive impacts of AI are equally compelling. In education, AI offers personalized learning platforms that adapt to individual needs, democratizing access to quality education. Healthcare also stands to benefit, as AI enhances capabilities in disease diagnosis, treatment planning, and patient monitoring, extending medical services to underserved regions.

AI's economic potential is transformative, driving innovation, creating new industries, and increasing productivity. It also contributes to urban planning, optimizing traffic flow, energy distribution, and public safety, thus fostering more sustainable and livable cities [23].

Balancing these dual impacts requires a thoughtful approach to AI development. Policymakers, developers, and stakeholders must collaborate to promote ethical AI practices, mitigate risks, and ensure equitable distribution of AI's benefits. By addressing these challenges proactively, society can harness AI's transformative potential while minimizing its adverse consequences [4].

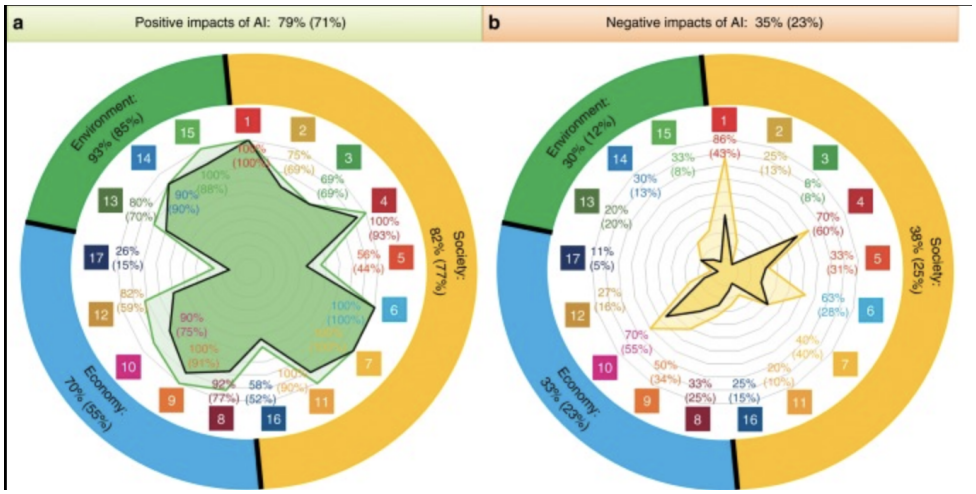


Figure 6.1: Summary of positive and negative impact of AI on the various SDGs. Documented evidence of the potential of AI acting as (a) an enabler or (b) an inhibitor on each of the SDGs. The numbers inside the colored squares represent each of the SDGs (see the Supplementary Data 1). The percentages on the top indicate the proportion of all targets potentially affected by AI and the ones in the inner circle of the figure correspond to proportions within each SDG. The results corresponding to the three main groups, namely Society, Economy, and Environment, are also shown in the outer circle of the figure. The results obtained when the type of evidence is taken into account are shown by the inner shaded area and the values in brackets. Source: [13]

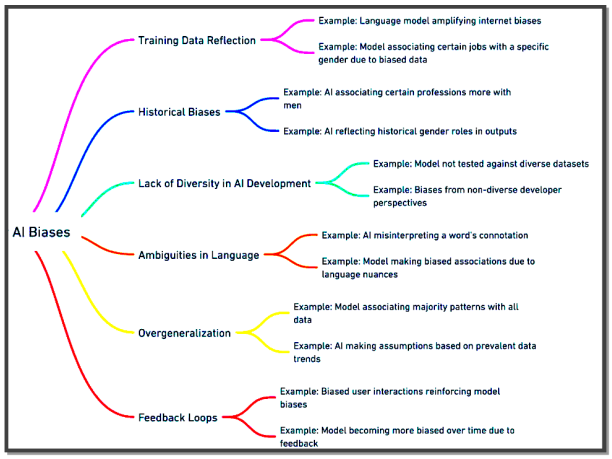


Figure 6.2: Mind map of AI Bias Sources and Examples

6.4.2 Root Causes

6.4.2.1 Biases in Algorithms

Algorithmic bias represents one of the most pressing challenges in AI deployment, manifesting through complex interconnected mechanisms that can perpetuate and amplify societal inequities. These biases arise from multiple pathways, each contributing to potentially discriminatory outcomes.

A fundamental source of algorithmic bias lies in the reflection of biases present in training data. Language models trained on internet data inadvertently absorb and amplify societal biases embedded in online discourse. For example, recruitment algorithms have been shown to disadvantage women by reproducing historical hiring patterns, prompting companies to abandon these tools after discovering their discriminatory outcomes [2].

Historical biases embedded in data further exacerbate systemic inequalities. Facial recognition systems, for instance, have exhibited significantly higher error rates for minorities, as seen in cases where misidentifications led to wrongful arrests.

The lack of diversity in AI development teams compounds these issues. Homogeneous development groups may fail to recognize biases that disadvantage underrepresented communities. This has been particularly evident in healthcare algorithms, which have demonstrated reduced accuracy for certain ethnic groups due to underrepresentation in training data [7].

Feedback loops present another challenge by amplifying biases over time. Social media algorithms, for example, can perpetuate echo chambers and polarization by recommending increasingly extreme content based on engagement patterns [4].

The broader consequences of these biases extend across critical domains. In healthcare, biased algorithms can exacerbate disparities in quality of care. In financial services, discriminatory credit scoring systems can limit economic opportunities for marginalized groups. Meanwhile, law enforcement applications of biased AI can result in disproportionate surveillance and policing of certain communities [12].

Addressing these challenges requires:

- **Regular bias audits**, incorporating sophisticated methodologies to identify and mitigate discriminatory patterns.
- **Diverse development teams** to ensure inclusive perspectives during the design and testing of AI systems.
- **Transparency requirements** that mandate clear documentation of training data sources and known biases .
- **Feedback mechanisms and monitoring systems** to continuously assess and correct biases in deployed systems.

As AI systems become increasingly integrated into societal functions, addressing biases is not just a technical necessity but a moral imperative. Fostering a comprehensive understanding of bias propagation will enable the development of more equitable and accountable AI systems.

6.4.2.2 Opaque Decision-Making

Opaque decision-making in artificial intelligence refers to the phenomenon where AI systems make decisions or predictions through processes that are difficult or impossible for humans to interpret, understand, or audit effectively. This opacity stems from both the intrinsic complexity of advanced algorithms and organizational practices that prioritize performance over transparency.

At the technical level, modern deep learning models, particularly those employing neural networks with millions or billions of parameters, operate through intricate transformations that defy straightforward human interpretation. The non-linear nature of these models means that even minor changes in input can produce dramatically different outputs through mechanisms that aren't easily traceable. Additionally, emergent properties in large language models can lead to behaviors that were neither explicitly programmed nor anticipated during training [16].

Organizational factors exacerbate this issue. Many companies treat their AI systems as "black boxes," citing intellectual property concerns to shield proprietary algorithms and training methodologies. This corporate opacity, combined with the technical complexity of AI systems, creates multiple layers of inscrutability. The competitive drive for higher performance often overshadows considerations of interpretability, further entrenching opaque practices.

Real-world examples underscore the implications of opaque decision-making. In the financial sector, AI-driven credit scoring systems have been criticized for their inability to provide clear reasons for decisions, such as the disparity in credit limits offered to individuals with similar profiles. Such opacity leaves consumers without recourse to challenge potentially unfair outcomes [5]. Similarly, in healthcare, diagnostic AI systems capable of remarkable accuracy often struggle to explain their reasoning, creating challenges for medical professionals who need to validate these recommendations.

The criminal justice system has also faced controversies with opaque AI applications, such as recidivism prediction tools. These algorithms, used to assess the likelihood of repeat offenses, often perpetuate historical biases. Without transparency, it becomes difficult to identify and rectify the sources of discriminatory patterns.

The societal impact of opaque decision-making extends beyond individual cases. When AI systems lack transparency, they erode public trust in technology and the institutions deploying it. This opacity can mask systemic discrimination and hinder effective oversight and regulation, allowing harmful practices to persist unchecked. Sustainable approaches to addressing opacity involve balancing proprietary interests with public demands for accountability and transparency.

Interpretable AI Development focuses on creating models that maintain high performance while offering clear explanations for their decisions. Techniques such as attention mechanisms in neural networks or inherently interpretable models can make AI systems more accessible and accountable [4].

Regulatory Frameworks like the European Union's GDPR establish a "right to explanation" for automated decisions, signaling a shift toward transparency in critical AI applications [6].

Hybrid Approaches combine complex AI systems with interpretable modules to offer explanations where they are most needed. These systems aim to balance the advantages of sophisticated algorithms with the necessity of human understanding [7].

Standardized Auditing Protocols are emerging to evaluate AI systems without requiring full transparency of proprietary algorithms. These protocols emphasize testing outcomes for fairness and identifying potential biases [3].

Sustainably addressing opaque decision-making demands a multifaceted approach, blending technical innovation with social responsibility. Progressive disclosure frameworks, regular impact assessments, and community engagement can foster trust and accountability while preserving the potential for innovation.

6.4.2.3 Economic and Access Inequities

The distribution of AI benefits and burdens across society reveals significant disparities in access and economic impact, with profound implications for global economic equality. A comprehensive analysis by PwC illustrates these stark disparities in expected AI-driven economic growth across different regions. The data projects that by 2030, China could see an extraordinary 26% GDP gain (US$7 trillion) from AI adoption, while North America expects a 14.5% increase (US$3.7 trillion). In contrast, regions like Africa, Oceania, and others are projected to see only a 5.6% gain (US$1.2 trillion), highlighting a concerning pattern of global inequality [1].

These projections reveal a potentially widening global digital divide that extends beyond mere technological access to encompass AI capabilities and their economic benefits. Developed regions, with their robust digital infrastructure, substantial R&D investments, and advanced technological expertise, are positioned to capture the majority of AI's economic advantages. This advantage creates a self-reinforcing cycle: regions with greater AI capabilities attract more investment, talent, and opportunities, further expanding their lead over less-developed areas.

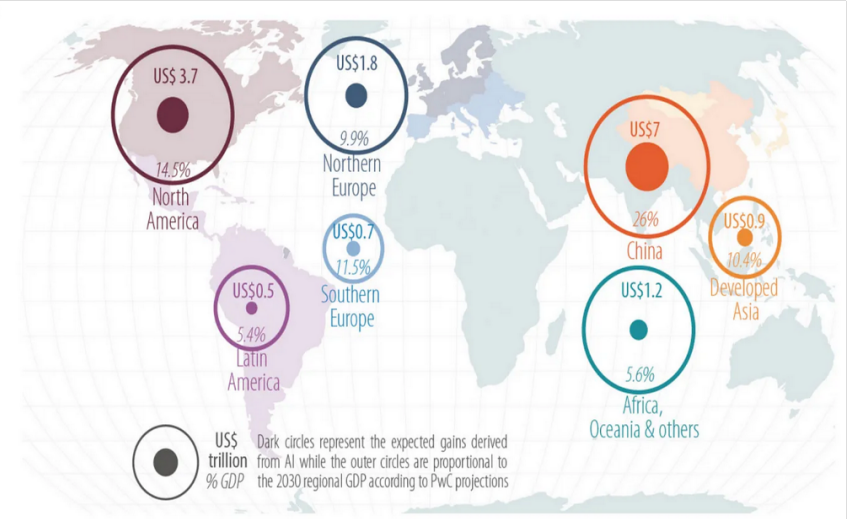


Figure 6.3: Regional disparities in projected AI-driven GDP growth by 2030. Dark circles represent expected gains from AI, while outer circles are proportional to regional GDP [13].

The implications of this disparity are particularly concerning for developing nations. While Northern Europe expects a 9.9% GDP gain (US$1.8 trillion) and Southern Europe projects 11.5% (US$0.7 trillion), Latin America anticipates only a 5.4% increase (US$0.5 trillion). This gap in AI-driven growth threatens to exacerbate existing global inequalities, potentially creating a new form of economic colonialism where AI capabilities determine a region's economic destiny.

The digital divide increasingly transforms into an AI divide, where communities with limited technological infrastructure or resources face exclusion from AI-driven services and opportunities. For example, automated customer service systems may be inaccessible to individuals without reliable internet access, while AI-powered educational tools might remain out of reach for underfunded schools [12]. The economic implications extend to job displacement, where workers in certain sectors face disproportionate risk of automation without adequate retraining opportunities [3].

These inequities manifest not only between nations but also within them. Even in developed countries, rural areas and underprivileged communities often lack access to AI-driven services and opportunities available in urban centers. This internal digital divide compounds existing socioeconomic disparities, creating multi-layered inequality that requires targeted intervention at both national and international levels [11].

Addressing these disparities requires a coordinated global response. International technology transfer programs must be established to help developing nations build AI capabilities, supported by substantial investment in digital infrastructure in underserved regions. These efforts should be accompanied by comprehensive capacity building initiatives to develop local AI talent and expertise. Furthermore, robust policy frameworks must be implemented to ensure AI benefits are distributed more equitably across society, preventing the concentration of technological advantages in already privileged regions and communities. The urgency of this response cannot be overstated, as delays in addressing these inequities risk cementing a new global hierarchy based on AI capabilities and access.

6.4.3 Regulatory Framework

The governance of artificial intelligence has evolved into a complex landscape of international regulations, national legislation, and global frameworks, each addressing different aspects of AI's societal impact. At the forefront of these regulatory efforts stands the General Data Protection Regulation (GDPR), which has fundamentally reshaped how

organizations handle automated decision-making processes. The GDPR's approach to AI governance is particularly significant in Article 22, which establishes crucial protections against purely automated decision-making that produces legal or similarly significant effects on individuals. This regulation mandates that individuals have the right to human intervention, to express their point of view, and to contest decisions made by AI systems. Organizations must provide clear explanations of the logic involved in automated decisions, ensuring transparency and accountability in AI-driven processes [6].

Article 22's implications extend far beyond simple data protection, creating a fundamental shift in how AI systems must be designed and deployed. For instance, financial institutions using AI for credit scoring must ensure their systems can provide comprehensible explanations for loan rejections, while employers using AI in recruitment must maintain human oversight in their hiring processes. The regulation also introduces the concept of "data protection by design and default," requiring organizations to embed privacy considerations into their AI systems from the earliest stages of development [6].

The European Union's AI Act represents the next evolution in AI regulation, introducing a comprehensive risk-based framework that categorizes AI applications based on their potential harm to society. This pioneering legislation establishes four risk levels: unacceptable risk (banned outright), high risk (subject to strict obligations), limited risk (requiring transparency), and minimal risk (permitted with minimal restrictions). High-risk applications, including critical infrastructure, educational assessment, law enforcement, and employment decisions, must meet stringent requirements for data quality, documentation, human oversight, accuracy, and robustness. The Act's extraterritorial scope means it affects any organization deploying AI systems that impact EU citizens, effectively setting global standards for AI development and deployment [10].

The United Nations Sustainable Development Goals (SDGs) provide a broader context for AI governance, particularly through Goals 9 (Industry, Innovation, and Infrastructure), 10 (Reduced Inequalities), and 16 (Peace, Justice, and Strong Institutions). While not specifically focused on AI, the SDGs offer a valuable framework for ensuring AI development aligns with global sustainability objectives. The integration of AI governance with SDGs highlights the importance of leveraging technological advancement to address global challenges while ensuring equitable access and preventing the exacerbation of existing inequalities [8].

The OECD AI Principles have emerged as a crucial international framework, establishing five complementary value-based principles for trustworthy AI. These principles emphasize inclusive growth, sustainable development, human-centered values, transparency, explainability, robustness, and accountability. Notably, the principles have been adopted by over 40 countries, including non-OECD members, demonstrating their global influence. The principles are particularly significant in their practical approach to AI governance, providing specific recommendations for national policies and international cooperation [12].

The U.S. National AI Initiative Act represents a different approach to AI governance, focusing on promoting innovation while ensuring ethical development and deployment. The Act establishes a coordinated federal strategy for AI research and development, emphasizing the importance of maintaining U.S. leadership in AI innovation while addressing societal concerns. Key provisions include the creation of the National AI Research Resource Task Force, which aims to democratize access to AI research tools and resources, and the emphasis on developing AI systems that promote equity and social welfare [5].

These regulatory frameworks interact in complex ways, creating a multi-layered governance structure. For instance, a multinational corporation developing AI systems must navigate GDPR requirements, comply with the EU AI Act's risk categories, align with OECD principles, and potentially meet U.S. regulatory requirements. This regulatory

complexity necessitates sophisticated compliance strategies and robust governance frameworks within organizations.

Implementation challenges include the need for standardized assessment methods for AI systems, mechanisms for ensuring meaningful human oversight, and procedures for demonstrating compliance with multiple regulatory regimes. Organizations must also balance competing requirements, such as the need for transparency in automated decision-making against the protection of proprietary algorithms and intellectual property.

Looking ahead, these regulatory frameworks continue to evolve. Emerging trends include increased focus on algorithmic auditing requirements, standardization of impact assessments, and development of certification schemes for AI systems. The challenge lies in maintaining regulatory effectiveness while fostering innovation and ensuring global coordination in AI governance [4].

The effectiveness of these regulatory frameworks depends significantly on enforcement mechanisms and international cooperation. While the GDPR has demonstrated the potential for substantial fines to ensure compliance, other frameworks rely more heavily on voluntary adoption and self-regulation. This diversity in enforcement approaches creates both challenges and opportunities for organizations developing and deploying AI systems.

6.4.4 Recommendations

Addressing the challenges of AI's social impact requires a multi-faceted approach combining technical solutions with policy interventions. Regular bias audits must become standard practice in AI development cycles, incorporating sophisticated testing methodologies to identify and eliminate discriminatory patterns in AI systems. These audits should extend beyond technical performance metrics to assess real-world impact on various demographic groups.

Policy support must evolve to keep pace with technological advancement, establishing clear accountability mechanisms and enforcement protocols. This includes developing standardized impact assessment frameworks and requiring regular compliance reviews for high-risk AI applications. Stakeholder engagement should be institutionalized through formal consultation processes that include affected communities, civil society organizations, and technical experts in policy development.

Transparency requirements must be strengthened, mandating explicability in AI decision-making processes, particularly in high-stakes domains like healthcare and criminal justice. Data privacy protections should be enhanced through state-of-the-art encryption protocols and robust access control mechanisms, with special attention to protecting vulnerable populations.

Finally, ethical training programs for AI developers should be mandatory and standardized across the industry, incorporating case studies of AI failures and successes in promoting social equity. These programs must emphasize practical approaches to bias mitigation and ethical decision-making in AI development.

6.5 Environmental Impact (Carbon Footprint of AI)

The difficulty of assessing the environmental impact of artificial intelligence has been covered previously. This chapter aims to introduce challenges and environmental impacts caused by artificial intelligence and showcase an overview of current regulations targeting the sustainability of artificial intelligence.

To give a detailed example of the environmental impact of AI, the carbon emissions and energy consumption caused by the inference and usage phase of artificial intelligence will be covered. Due to the aforementioned complexity of measuring the environmental impact, we will focus only on the direct energy consumption and carbon emissions and omit the grey energy used to create the hardware, as well as the further environmental impacts. This is done to allow a deeper analysis, be concise, and avoid touching on a multitude of subjects superficially. However, we do not forget the multifaceted nature of the environmental implications of AI, especially since they are relevant for regulations and possible future regulations.

6.5.1 Energy Consumption and Emissions

Artificial intelligence systems, especially large-scale models, have become central to technological innovation. Funding for AI development has increased by a factor of over 7 since 2015, reaching $93.5 billion in 2021, and has continued to grow since then [56].

Large amounts of energy are required for the inference phase of AI models. The training of GPT-3 consumed approximately 1,287 megawatt-hours (MWh) of electricity, resulting in over 500 tons of CO_2 emissions [54, p.3]. This is comparable to the annual emissions of dozens of average households. These figures underscore the significant energy footprint of training AI models. GPT-3, now a four-year-old model, is a case in point. Since its development, the number of AI models being trained has increased substantially, exacerbating energy consumption due to the scalability of the issue. Additionally, energy consumption for model training has been doubling approximately every 3.4 months, according to Clemm et al. For instance, the transition from GPT-2 to GPT-3 resulted in a 20-fold increase in computational resources, directly translating to higher energy usage [57, p.5]. The combination of the growing number of AI models and the increasing energy demands for training each new iteration raises critical questions about the environmental sustainability of AI.

The global energy mix remains dominated by non-renewable sources, particularly in countries most involved in the development of artificial intelligence, such as the United States, China, Japan, and nations within the European Union [55]. This means that the training and use of AI systems not only account for significant electricity consumption but also result in substantial carbon emissions, thereby contributing to global warming.

Efforts to improve energy efficiency in AI have yielded mixed results, with the significant energy consumption of AI risking the offsetting of its potential environmental benefits. While AI-powered optimization has been shown to improve energy efficiency in industrial processes, the energy savings facilitated by AI must outweigh the energy required for model training to achieve true environmental sustainability [54, p.3, 11].

6.5.2 Existing Regulations

To provide a clearer overview of current regulations targeting the increase in the sustainability of AI, they will be assigned to three categories depending on their level of abstraction and generality. Namely, from higher to lower levels. On the highest level are regulations or goals agreed upon by international organizations or contracts signed by the vast majority of countries. These agreements tend to be general. The second category consists of regulations mandated by individual countries, or supranational unions like the European Union, which are more specific to AI and target AI directly.

The third category will cover regulations or practices that are either voluntary or indirectly influence the sustainability of AI.

Regulations themselves are a rather broad term that can either be specific if defined or be understood as a broader term for anything that aims to regulate something, which could be laws, but also contracts between individual countries, private parties, etc. For example, the European Commission has multiple categories of EU laws, of which regulations are one category [48]. In this chapter, we will use the broader term of regulations, meaning we will also encompass directives, recommendations, and any other regulatory forms targeting the environmental sustainability of AI. We will specify if a recommendation of a certain body or even contracts between industry partners are voluntary.

Prior to covering specific regulations, a major difficulty for researchers and lawmakers will be covered. To be able to regulate AI and its sustainability, a definition for artificial intelligence is required. Lawmakers and researchers are faced with the challenge of not having a consensus regarding a definition of artificial intelligence and, additionally, possible definitions themselves evolving due to the evolution of AI itself [41]. A study by the European Parliamentary Research Service concludes that defining AI is a nearly impossible task and that said ambiguity reflects on the Act said study critiques [42, p. 10].

This challenge is not unique to said Act and provides challenges for all regulators and researchers [41]. The study does reference a definition of AI as computer programs that emulate human, rational behavior, and thinking [42, p. 10]. The Regulation 2024/1689 of the European Parliament and Council, commonly known as the EU Artificial Intelligence Act, however, defines an AI system as a "machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments" [43, p. 46]. This definition is inherently different from the one used in the study [42] and this comparison should showcase the difficulty that AI regulations bring with them before even being able to go into the specifics. To focus on the regulations targeting AI sustainability, the definition of AI in each regulation, if present, will not be covered or discussed.

Regulations influencing the environmental sustainability of AI on the highest regulatory level include the Paris Agreement. The Paris Agreement, being the first binding agreement adopted by 196 parties to address climate change and keep global warming below 2°C [44], aims to maximize the benefits of local and regional energy transitions. A supportive policy framework is crucial to address climate impacts. Global agreements like the Paris Agreement can play a key role in guiding green initiatives, incentivizing shared green investments, technologies, and building green energy capacity across nations [45, p. 3-4]. AI, being a shareable technology and offering great potential in reducing energy con-

sumption through optimization [46], can, by embedding it within clean energy initiatives supported by international collaboration under the Paris Agreement, contribute to the sustainable use of artificial intelligence. A global agreement such as the Paris Agreement provides stability for consistent progress for sustainable AI-driven innovations by mitigating geopolitical instability [45, p. 4-5]. However, by focusing on zero-carbon solutions and promoting green technologies, AI itself will be either directly or indirectly influenced by policies and frameworks established through the Paris Agreement due to the previously showcased substantial energy consumption of AI, especially in the training phase and the emissions caused.

Regulations targeting the environmental sustainability in the second category include the already covered EU AI Act. While mainly focusing on the ethical and social sustainability of artificial intelligence, the EU AI Act encourages transparency regarding AI lifecycle energy consumption and environmental sustainability. The Act requires risk management processes that include identifying potential environmental impacts for AI systems to be incorporated [10]. The Act's documentation and monitoring requirements align with methodologies like the life cycle assessment. Additionally, by encouraging the minimization of risks and improving transparency in AI development and use, the Act indirectly promotes energy-efficient models and the reduction of carbon emissions.

The Corporate Sustainability Reporting Directive (CSRD) similarly impacts the environmental sustainability of AI indirectly by requiring companies to report their sustainability efforts and implement an extensive reporting and auditing framework [49]. This transparency addresses the current challenges of lacking data for assessing the environmental impact of artificial intelligence. Furthermore, the publication of such data can be expected to incentivize companies to adopt more sustainable practices.

The third category of regulations consists of those not directly targeting AI but indirectly impacting its sustainability. These regulations are mainly frameworks applying to data centers as well as industry standards for energy efficiency or recycling. Data centers and general ICT systems are the vessels on which artificial intelligence is run and operated. In conclusion, any regulations targeting the environmental sustainability of said systems indirectly target the environmental sustainability of artificial intelligence as well.

The EU Energy Efficiency Directive requires large enterprises, including data centers, to undergo energy audits and implement energy management systems to enhance efficiency [50]. The ISO/IEC 30134 Series provides metrics for data center resource efficiency, including PUE, aiding in the assessment and improvement of energy performance and can be sorted into the third category [51].

Similarly, the environmental sustainability of AI is indirectly impacted by a number of regulations targeting either the ethical or social sustainability or data centers in China. The Action Plan for Green Development of Data Centers (July 2024) focuses on improving the energy efficiency of data centers, especially in AI-heavy sectors. It aims to reduce Power Usage Effectiveness (PUE) and increase the use of renewable energy sources. Optimizing the geographical locations of data centers in China aims to mitigate environmental impacts [52, p. 290].

While primarily addressing ethical concerns, the Deep Synthesis Regulation (effective January 2023) emphasizes responsible AI usage and indirectly addresses energy sustainability [52, p. 292-293].

The Algorithm Filing Requirements (effective June 2024) encourage the optimization of algorithms for efficiency by requiring detailed reports on their functions, impacts, and mitigation strategies. Higher algorithmic efficiency leads to reduced energy consumption, thereby further targeting the environmental sustainability of artificial intelligence [52, p. 292-294].

6.5.3 Recommendations: Green AI Practices, Carbon Offsetting

Lastly, we aim to introduce possible approaches covered by researchers to make AI more environmentally sustainable as well as our own thoughts. We want to mention the rather dynamic legislative landscape surrounding artificial intelligence as well as the rapidly evolving industry aiming to make AI sustainable. Proposals may intersect with already proposed or even implemented regulations, recommendations, or laws. However, we would like to present our key thoughts regarding possible approaches and focuses.

1. **Establish AI Carbon Reporting Standards** To capture the effective impact of AI, extensive data regarding emissions and energy consumption needs to be available. Many regulations already incentivize more transparency. However, much of the reporting is done on a higher company level and not narrowed down to individual technologies, meaning there is still room for more concrete data.

2. **Adopt Life Cycle Thinking for AI Systems** Verdecchia et al. advocate for employing a holistic approach that considers all stages of an AI system's life cycle, including hardware manufacturing, data storage, training, and inference. Aiming to ensure that measures to increase sustainability are integrated into each step, from model design to deployment [53, p. 4].

3. **Deploying Energy-Efficient Models** Wu et al. analyze not just the substantial energy consumption of LLMs but also showcase possible rewards of more efficient models [54, p. 4-6]. Further optimizing models to reduce their energy consumption will increase environmental sustainability.

4. **Incentivize Use of Renewable Energy** Verdecchia et al. propose transitioning data centers to renewable energy sources and optimizing their geographic locations to regions with cleaner energy to reduce the carbon footprint of AI infrastructure [53, p. 10-12]. Additionally, the use of solely renewable energy could be incentivized by tax breaks for data centers purely operating on renewable energy.

6.6 Ethical Implications of AI

6.6.1 Distinguishing Social and Ethical Impacts of AI

While interrelated, the social and ethical impacts of AI represent distinct dimensions of technological influence on human society. Social impacts encompass the tangible, observable effects of AI systems on human communities, interactions, and societal structures. These manifest in concrete ways: the transformation of workplaces through automation leading to job displacement and role redefinition; the reshaping of human interactions through AI-mediated communication platforms; the amplification or mitigation of existing social inequalities through algorithmic decision-making; and the varying levels of technology accessibility across different demographic groups. For instance, AI-driven recruitment tools directly affect employment opportunities in specific communities, while automated content moderation systems actively shape public discourse and social connections [2].

In contrast, ethical impacts concern the fundamental moral principles and values that guide AI development and deployment. These impacts focus on universal considerations that transcend specific social contexts: the imperative for fairness in algorithmic decision-making; the requirement for transparency in AI systems; the necessity of accountability in automated processes; and the fundamental obligation to prevent harm across all applications [7]. Unlike social impacts, which can be observed in specific communities, ethical impacts deal with overarching principles that apply universally, regardless of the social context or affected population. For example, while the social impact of an AI healthcare diagnostic tool might be measured in terms of patient outcomes in specific communities, its ethical impact concerns broader questions about patient autonomy, informed consent, and the fair distribution of healthcare resources [3].

The distinction becomes particularly relevant when addressing AI governance: social impacts often drive specific policy interventions and mitigation strategies for particular communities, while ethical impacts inform the fundamental principles and values that should guide AI development across all contexts and applications. Understanding this distinction enables more effective approaches to both addressing immediate societal challenges and ensuring long-term responsible AI development aligned with human values and moral principles [9; 1].

6.6.2 Ethical AI: Principles and Impact

Ethical AI represents a critical framework for developing and deploying artificial intelligence systems in ways that align with human values, promote societal wellbeing, and prevent harm. As defined by leading institutions, AI ethics encompasses a set of moral principles that guide the development and implementation of AI technologies, focusing on optimizing beneficial impacts while minimizing potential risks and adverse outcomes [7]. This multidisciplinary field draws from philosophy, computer science, sociology, and other domains to create comprehensive guidelines for responsible AI development.

The importance of ethical AI extends far beyond mere legal compliance, establishing proactive principles that anticipate and address potential challenges before they manifest as societal problems. While legal frameworks often lag behind technological advancement, ethical guidelines serve as a dynamic compass for innovation, encouraging developers and organizations to consider the broader implications of their work [9]. This proactive approach helps prevent the deployment of AI systems that, while technically legal, might still cause unintended harm or violate ethical principles.

Human rights protection stands as a fundamental pillar of ethical AI development. This includes safeguarding privacy rights, preventing discriminatory outcomes, and ensuring algorithmic fairness across diverse populations. For instance, facial recognition sys-

tems must be developed with careful consideration of privacy implications and potential misuse for surveillance. Similarly, AI-driven hiring systems must be designed to prevent discrimination based on protected characteristics while promoting fair opportunity for all candidates [2].

The relationship between ethical AI and public trust cannot be overstated. As AI systems increasingly influence critical aspects of human life—from healthcare decisions to financial opportunities—maintaining public confidence becomes essential for widespread adoption and effective implementation. Organizations that demonstrate strong commitment to ethical AI principles often find greater acceptance of their technologies among users and stakeholders. This trust becomes particularly crucial in sensitive applications like medical diagnosis or autonomous vehicle operation, where public confidence directly impacts the technology's potential benefits to society [3].

Responsible innovation in AI development requires careful consideration of long-term consequences and societal impact. Ethical AI frameworks encourage developers to consider not just technical capabilities but also social responsibility, environmental sustainability, and cultural implications. This includes addressing questions about AI's impact on employment, social relationships, and human agency. For instance, the development of AI-driven automation must balance efficiency gains against potential workforce displacement, considering ways to create new opportunities while mitigating negative impacts on affected communities [1; 4].

The implementation of ethical AI principles requires systematic approaches across multiple dimensions. Organizations must establish clear governance structures that incorporate ethical considerations at every stage of AI development and deployment. This includes diverse representation in development teams, regular ethical impact assessments, and transparent communication about AI capabilities and limitations. Regular auditing and monitoring ensure that AI systems continue to meet ethical standards as they evolve and interact with real-world scenarios [7; 10].

Sustainability in ethical AI development demands consideration of both immediate and long-term impacts. This includes environmental sustainability through efficient resource use and reduced energy consumption, as well as social sustainability through fair access to AI benefits across different societal groups. Economic sustainability requires balancing innovation with responsible development practices that create lasting value while preventing harmful disruptions to communities and industries [2; 1].

The global nature of AI deployment necessitates consideration of cultural differences and varying ethical frameworks across societies. What might be considered ethical in one context could raise concerns in another, requiring flexible and culturally sensitive approaches to AI development and deployment. This cultural awareness becomes particularly important as AI systems are deployed across international borders and diverse communities [9; 3].

Educational initiatives play a crucial role in promoting ethical AI development and implementation. This includes training for AI developers in ethical principles and implications, as well as broader public education about AI capabilities, limitations, and potential impacts. Informed stakeholders can better participate in discussions about AI governance and help shape the development of AI systems that serve societal needs while respecting ethical boundaries [7; 10].

As AI technology continues to advance and integrate more deeply into society, the importance of ethical AI principles grows increasingly evident. These principles serve not as constraints on innovation but as guidelines for developing AI systems that create lasting positive impact while minimizing potential harms. Through careful attention to ethical considerations, the AI industry can build technologies that not only advance human capabilities but also protect and promote human values and wellbeing [3; 1].

6.6.3 Transparency and Explainability

A significant ethical concern in AI revolves around the "black box" nature of many AI systems, especially those utilizing complex algorithms like deep learning. These models often operate with levels of abstraction that make their decision-making processes opaque, even to their developers. In high-stakes applications such as credit scoring, healthcare diagnostics, and criminal justice, this lack of transparency poses severe ethical dilemmas, as affected individuals and stakeholders cannot fully understand, challenge, or rectify decisions that impact their lives.

Ethical standards emphasize the importance of explainability to ensure that AI decisions are not only accurate but also interpretable. Explainability provides a means for affected parties to contest decisions, which is particularly important in applications that influence fundamental rights. The European Union's AI Act mandates transparency obligations for high-risk AI systems, requiring that users be informed when interacting with AI and given understandable explanations for critical decisions [10]. Similarly, the Ethics Guidelines for Trustworthy AI advocate for transparency as a pillar of ethical AI, insisting that systems should provide clear insights into their operations to foster trust among users and regulators alike [7]. By mandating explainability, these frameworks aim to reduce the opacity that currently characterizes many AI systems, ensuring that ethical obligations are upheld across sectors.

6.6.4 Accountability and Responsibility

As AI systems increasingly assume roles traditionally held by humans, determining accountability in cases of harm, error, or bias becomes a pressing ethical challenge. Accountability frameworks in AI seek to assign responsibility to various stakeholders—developers, data providers, and operators—each of whom plays a role in the system's design, deployment, and operation. Establishing clear accountability pathways is essential to address ethical lapses and ensure that affected individuals have recourse in cases where AI-driven decisions result in harm.

The ethical principle of accountability requires that developers and organizations anticipate the potential impacts of their systems and implement measures to prevent harm. The AI Act introduces accountability requirements for high-risk AI applications, such as mandatory risk assessments, documentation, and audits, to ensure that developers can be held responsible for their system's outputs [10]. Additionally, industry standards like those proposed by AI4People underscore the ethical need for companies to establish internal accountability structures, such as ethics committees or oversight boards, that can scrutinize and evaluate the social and ethical implications of AI systems [9]. By embedding accountability within the governance of AI, these frameworks aim to uphold ethical standards and ensure that AI systems are deployed in ways that respect the rights and welfare of all users.

6.6.4.1 Distinction between Legal AI and Ethical AI

The distinction between legal and ethical AI frameworks is crucial in the development and deployment of artificial intelligence systems. Although these frameworks often overlap, they serve distinct purposes, addressing different dimensions of responsibility and governance. Legal frameworks establish enforceable standards, while ethical frameworks provide broader guidelines that reflect societal values and moral principles [6; 7; 10].

Legal AI frameworks are codified regulations that define minimum requirements for AI development and deployment. Typically emerging from legislative processes, these frameworks impose specific, measurable criteria to ensure compliance and enforce penalties for violations. For instance, the European Union's General Data Protection Regulation

(GDPR) mandates explicit consent for the processing of personal data and grants individuals the right to explanation for automated decisions [6]. Similarly, the Illinois Biometric Information Privacy Act (BIPA) regulates the collection and use of biometric data, directly influencing the design and application of facial recognition systems.

The primary aim of legal frameworks is to establish clear boundaries that protect fundamental rights and ensure fairness, privacy, and non-discrimination. For example, data privacy laws like the California Consumer Privacy Act and anti-discrimination mandates in employment law set enforceable standards to safeguard individuals and communities. However, these frameworks often struggle to keep pace with rapid technological advancements, leaving regulatory gaps that fail to address emerging issues such as subtle algorithmic biases or the unintended consequences of complex AI interactions. Additionally, the specificity of legal mandates can lead to rigid requirements that overlook nuanced ethical considerations.

In contrast, ethical AI frameworks transcend the legal minimum to emphasize proactive responsibility, long-term societal impact, and alignment with human values. Ethical frameworks guide organizations in addressing broader questions, such as equitable access to AI technologies, the preservation of human agency, and the responsible development of increasingly autonomous systems. These frameworks encourage developers and organizations to consider the societal consequences of their innovations, promoting fairness, accountability, and sustainability in AI applications.

The distinction between legal and ethical frameworks becomes evident in real-world scenarios. For instance, an AI-driven hiring system might comply with legal requirements by avoiding explicit discrimination based on protected characteristics like race or gender. However, ethical considerations would require evaluating the system's broader impact on workplace diversity, human agency in hiring decisions, and the potential reinforcement of societal inequities. Similarly, in healthcare, legal frameworks might protect patient privacy and ensure the security of medical records. Ethical frameworks, however, extend these considerations to include the balance between AI-driven decision-making and human judgment, equitable access to AI-enhanced healthcare, and the preservation of the doctor-patient relationship.

Navigating the intersection of legal and ethical frameworks poses challenges and opportunities. Organizations must simultaneously comply with legal mandates and address broader ethical considerations, which often requires the establishment of governance structures capable of adapting to evolving requirements. Such governance must ensure compliance across jurisdictions while maintaining consistent ethical standards globally. Moreover, ethical challenges not covered by existing legal frameworks, such as the responsible use of advanced AI capabilities, demand forward-thinking approaches that prioritize societal well-being.

The evolving relationship between legal and ethical frameworks highlights their interdependence. Ethical principles often inspire the creation of new legal standards, while legal frameworks provide a baseline for operationalizing ethical goals. This dynamic underscores the importance of viewing legal compliance as a foundation rather than an endpoint, encouraging organizations to strive for higher ethical standards that foster responsible innovation and sustainable development [6].

Achieving this balance requires an ongoing dialogue among technologists, ethicists, legal experts, and stakeholders impacted by AI systems. By aligning legal and ethical frameworks, organizations can ensure that AI technologies advance human welfare, uphold societal values, and address emerging challenges effectively. The ultimate objective is to create AI systems that not only meet regulatory requirements but also contribute to a more equitable and sustainable future.

6.6.5 Recommendations: Ethical AI Frameworks

To address the ethical challenges posed by AI, it is necessary to establish comprehensive ethical AI frameworks that prioritize transparency, bias mitigation, and accountability. These frameworks should be developed through a collaborative approach involving governments, industry leaders, civil society, and academia to ensure they are robust, enforceable, and adaptable to the rapidly evolving nature of AI technologies. Furthermore, these frameworks must be both global and context-sensitive, accounting for variations in legal, cultural, and societal expectations across different regions while aligning with universal ethical principles.

6.6.5.1 Bias Mitigation as an Ethical Duty

While the social impact of bias addresses the specific effects on marginalized groups, bias mitigation as an ethical duty speaks to the overarching responsibility of developers and companies to identify and reduce bias in AI systems proactively. Ethical guidelines emphasize that AI developers have a duty to detect and address biases that may exist within datasets or algorithms, preventing the reinforcement of harmful stereotypes and inequities. Regular bias audits and bias mitigation strategies are essential components of ethical AI, ensuring that systems operate fairly and that any biases are minimized before deployment.

The AI4People framework and the Trustworthy AI guidelines stress that bias mitigation should be integrated into every stage of the AI lifecycle, from data collection and model design to deployment and monitoring [9; 7]. This proactive approach helps to safeguard against discriminatory practices and fosters greater equity in AI applications. Ethical AI mandates that developers consider the potential for harm and actively work to ensure that their systems do not perpetuate unfair biases. By treating bias mitigation as an ethical obligation, AI governance frameworks can promote fairness, reliability, and societal trust in AI systems.

6.6.5.2 Alignment with Ethical Guidelines and Standards

A robust ethical framework for AI requires adherence to recognized guidelines and standards, which establish universal principles for the responsible development of AI technologies. The European Union's AI Act, GDPR, and AI4People's ethical principles each provide a structured approach for embedding ethical practices within AI development. These guidelines advocate for transparency, accountability, fairness, and respect for fundamental rights, creating a regulatory environment that promotes ethical AI across various applications.

AI developers and organizations are encouraged to adopt ethical frameworks, such as the Trustworthy AI guidelines, to guide decision-making throughout the AI lifecycle [7]. These frameworks establish best practices for data use, model design, and user interaction, ensuring that AI systems operate in alignment with societal values and legal norms. Ethical guidelines not only foster responsible AI but also offer organizations a structured pathway for compliance, accountability, and public trust. By aligning AI systems with these standards, developers and policymakers can create a safer, more ethical AI ecosystem that respects human rights and mitigates potential harms.

6.6.5.3 Ongoing Research and Ethical Education

Promoting ethical AI requires sustained investment in research and education. Governments, academic institutions, and private organizations should invest in research that explores the ethical, social, and legal implications of AI technologies. This research should

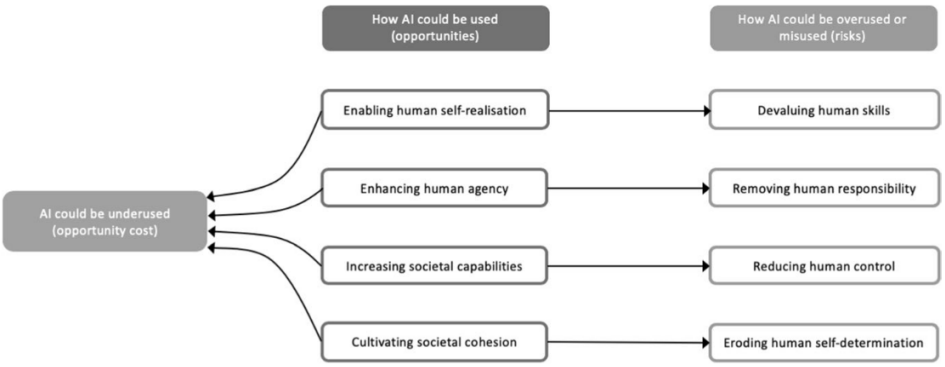


Figure 6.4: Overview of the four core opportunities offered by AI, four corresponding risks, and the opportunity cost of underusing AI. Source: [9]

not only focus on technical solutions, such as improving the fairness and transparency of AI systems, but also on the broader societal impacts of AI adoption, such as labor displacement, surveillance, and privacy concerns.

As illustrated in Figure 6.4, AI has immense potential to enhance human capabilities, societal cohesion, and agency. However, overuse or misuse of AI risks undermining these very benefits. For example, while AI can cultivate societal cohesion and improve decision-making, excessive reliance on it may erode human self-determination and diminish essential skills. This duality underscores the need for ethical education and governance that emphasize responsible use while mitigating risks.

Interdisciplinary research is particularly important for understanding these ethical challenges. While computer scientists are crucial for developing technical solutions, insights from fields such as sociology, law, philosophy, and political science are necessary to address the complex ethical and social questions surrounding AI deployment. Collaborative research efforts that bring together these diverse perspectives are more likely to result in comprehensive ethical frameworks that can be effectively applied in practice [12].

In addition to research, AI ethics education should be integrated into computer science curricula at all levels of higher education. As AI technologies continue to advance, it is crucial that the next generation of AI developers and engineers are equipped with the knowledge and tools to build ethical systems. Universities should offer courses that cover topics such as algorithmic bias, data ethics, AI governance, and the societal impacts of AI. Moreover, ethics should be embedded throughout computer science education, ensuring that students understand the ethical implications of their work from the very beginning of their training.

Public awareness campaigns are also essential for empowering individuals to protect their rights when interacting with AI systems. Governments and organizations should launch campaigns that educate the public about the ethical implications of AI, how AI systems work, and what steps individuals can take to contest unfair or harmful decisions made by AI. By fostering a more informed public, these campaigns can help build trust in AI technologies while also promoting greater accountability for AI developers and operators

6.6.5.4 Global Collaboration and Universal Standards

Finally, there is an urgent need for global collaboration in developing ethical AI standards. AI is a global technology, and its ethical challenges—ranging from bias and surveillance to labor displacement and environmental impact—transcend national borders. International organizations, such as the United Nations, the European Union, and the World Economic Forum, should lead efforts to create a set of universal ethical AI standards that can

be adopted by countries worldwide. These standards should emphasize transparency, accountability, fairness, and inclusivity, ensuring that AI systems are deployed in ways that align with shared human values and contribute to the common good [11].

A globally coordinated approach to ethical AI development is essential for addressing cross-border issues, such as the global flow of data, the deployment of AI in international finance and trade, and the use of AI in surveillance and defense systems. By establishing harmonized regulatory frameworks and ethical standards, countries can ensure that AI systems are developed in accordance with international human rights laws and sustainable development goals.

Moreover, global collaboration should involve capacity-building initiatives to help developing countries adopt ethical AI practices. Many low- and middle-income countries are currently left out of discussions on AI governance, despite the fact that they will be significantly affected by the deployment of AI systems, particularly in areas like agriculture, healthcare, and education. By providing technical and regulatory support to these countries, the global community can help ensure that AI is used to promote equitable and sustainable development worldwide.

6.7 Summary and General Recommendations

6.7.1 Key Findings

This report examined the multifaceted relationship between artificial intelligence (AI) and sustainability, encompassing environmental, social, and ethical dimensions. It highlights that AI, while holding immense potential, also presents significant risks that require management and governance.

Key findings include the challenges in assessing AI's environmental impact due to data scarcity and system complexity. Finding metrics covering all three sustainability dimensions was shown to be unfeasible in the current state, highlighting the need for further research. The significant energy consumption associated with training and deploying large-scale AI models further poses a risk of outweighing possible benefits. Socially, AI systems risk perpetuating biases, exacerbating inequities, and concentrating technological benefits in a limited number of regions or groups, further increasing economic and social inequality. Ethical concerns, including opacity, accountability gaps, and uneven adherence to ethical guidelines, further complicate AI's integration into society. The black-box nature of AI, particularly large language models (LLMs), hinders the analysis and understanding of these systems.

The analysis also highlighted existing regulatory frameworks, such as the EU AI Act, General Data Protection Regulation (GDPR), and data center standards, which address some of these issues but leave gaps in enforceability and adaptability due to the pace of rapidly evolving technologies in a highly competitive market. Global frameworks like the United Nations Sustainable Development Goals provide overarching principles but lack specificity in their application to AI.

6.7.2 Proposed Guidelines for Future Regulatory Frameworks

To address the challenge of balancing the opportunities and risks posed by artificial intelligence (AI), a comprehensive, adaptive, and forward-looking approach to regulation is required. The main guidelines and concepts developed in this report can be structured as follows:

1. **Integrated Lifecycle Assessments**: Regulations should mandate the adoption of lifecycle assessments for AI systems, to capture the full environmental Impact from

hardware production, training, deployment, up to the decommissioning phases. This would facilitate a comprehensive understanding of environmental impacts and guide sustainable practices.

2. **Enhanced Transparency Requirements**: Future frameworks must enforce transparency in AI systems, particularly for high-risk applications. This includes explainability in decision-making processes, clear documentation of data sources, and mandatory disclosure of environmental footprints.

3. **Ethical Standards as Regulatory Baselines**: Ethical AI principles, such as fairness, accountability, and inclusivity, should be embedded as regulatory baselines. This ensures that AI development aligns with societal values and mitigates risks of harm or discrimination.

4. **Promoting Global Collaboration**: Given AI's global impact, international cooperation is crucial. Regulatory frameworks should encourage the harmonization of standards, facilitate knowledge sharing, and establish mechanisms for joint enforcement across jurisdictions.

5. **Focus on Social Equity and Access**: Regulations must prioritize equitable access to AI benefits. This includes incentivizing AI deployment in underserved regions, supporting education and capacity building, and addressing the digital divide.

6. **Dynamic and Adaptive Policies**: Recognizing the rapid evolution of AI, regulatory frameworks should incorporate mechanisms for periodic review and adaptation. This ensures continued relevance and effectiveness of legal frameworks.

7. **Incentivizing Green AI**: Governments should establish incentives for developing and deploying energy-efficient AI models and transitioning to renewable energy sources. Tax breaks, grants, and public-private partnerships can encourage adherence to sustainable practices.

8. **Strengthened Accountability Mechanisms**: Clear pathways for accountability should be established, detailing responsibilities for developers, operators, and regulators. Regular audits and compliance checks must be standard practice for high-risk AI systems.

Bibliography

[1] F. Rohde et al.: Broadening the perspective for sustainable artificial intelligence: Sustainability criteria and indicators for artificial intelligence systems; Journal article (Current Opinion in Environmental Sustainability, Vol. 66), 2024, https://doi.org/10.1016/j.cosust.2023.101411.

[2] M. A. Goralski and T. K. Tan: Artificial intelligence and sustainable development; Journal article (The International Journal of Management Education, Vol. 18, No. 100330), 2020, https://doi.org/10.1016/j.ijme.2019.100330.

[3] A. van Wynsberghe: Sustainable AI: AI for sustainability and the sustainability of AI; Journal article (AI and Ethics, Vol. 1, pp. 213–218), 2021, https://doi.org/10.1007/s43681-021-00043-6.

[4] P. Hacker: Sustainable AI Regulation; Journal article (Common Market Law Review, forthcoming), 2024, https://arxiv.org/abs/2306.00292.

[5] R. Dhiman et al.: Artificial intelligence and sustainability: A review; Journal article (Analytics, Vol. 3, pp. 140–164), 2024, https://doi.org/10.3390/analytics3010008.

[6] European Union: General Data Protection Regulation (GDPR); Regulation (EU) 2016/679, 2018.

[7] European Commission: Ethics guidelines for trustworthy AI; Report (Brussels, Belgium), 2019.

[8] United Nations: Sustainable Development Goals (SDGs); Global framework, 2015.

[9] L. Floridi et al.: AI4People – An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations; Journal article (Minds and Machines, Vol. 28, pp. 689–707), 2018, https://doi.org/10.1007/s11023-018-9482-5.

[10] European Commission: Artificial Intelligence Act (AI Act); Regulation, 2021.

[11] McKinsey Digital: AI for social good: Improving lives and protecting the planet; Report, 2024.

[12] S. Larsson et al.: Sustainable AI: An inventory of the state of knowledge of ethical, social, and legal challenges related to artificial intelligence; Report (AI Sustainability Center, Lund University), 2019.

[13] R. Vinuesa et al.: The Role of Artificial Intelligence in Achieving the Sustainable Development Goals; Journal article (Nature Communications, Vol. 11, pp. 1–10), 2020.

[14] C. Freitag et al.: The Real Climate and Transformative Impact of ICT: A Critique of Estimates, Trends, and Blind Spots; Journal article (Patterns, Vol. 2, No. 9, pp. 100340), 2021, https://doi.org/10.1016/j.patterns.2021.100340.

[15] D. Patterson et al.: Carbon Emissions and Large Language Models; Journal article (Communications of the ACM, Vol. 65, No. 12, pp. 54–65), 2022.

[16] R. Bommasani et al.: On the Opportunities and Risks of Foundation Models; Preprint (arXiv), 2021, https://arxiv.org/abs/2108.07258.

[17] G. Gayatri Tanuja et al.: Innovative Technologies for Sustainable Recycling and Re-manufacturing of Materials and Components; Conference paper (E3S Web of Conferences, Vol. 430, pp. 01130), 2023, https://doi.org/10.1051/e3sconf/202343001130.

[18] A. Lacoste et al.: Tackling Climate Change with Machine Learning; Journal article (Nature Climate Change, Vol. 9, No. 8, pp. 629–633), 2019.

[19] T. Hagendorff: The Ethics of AI Ethics: An Evaluation of Guidelines; Journal article (Minds and Machines, Vol. 30, pp. 99–120), 2020, https://doi.org/10.1007/s11023-020-09517-8.

[20] A. Jobin et al.: The global landscape of AI ethics guidelines; Journal article (Nature Machine Intelligence, Vol. 1, pp. 389–399), 2019, https://doi.org/10.1038/s42256-019-0088-2.

[21] D. Rolnick et al.: Tackling Climate Change with Machine Learning; Preprint (arXiv), 2019, https://arxiv.org/abs/1906.05433.

[22] N. A. Christakis: We need to focus more on the social effects of AI, says Nicholas Christakis; Online article (*The Economist*), December 2023, https://www.economist.com/by-invitation/2023/12/15/we-need-to-focus-more-on-the-social-effects-of-ai-says-nicholas-christakis, Accessed: November 27, 2024.

[23] F. L. Ruta: Do the Benefits of Artificial Intelligence Outweigh the Risks?; Online article (*The Economist: Open Future*), September 2018, https://www.economist.com/open-future/2018/09/10/do-the-benefits-of-artificial-intelligence-outweigh-the-risks, Accessed: November 27, 2024.

[24] R. Torelli: Sustainability, responsibility and ethics: different concepts for a single path; Journal article (Social Responsibility Journal, Vol. 17, No. 5, pp. 719–739), 2021, https://doi.org/10.1108/srj-03-2020-0081.

[25] ADEC Innovations: What is Social Sustainability?; Online resource, 2022, https://www.adecesg.com/resources/faq/what-is-social-sustainability/.

[26] V. Levesque: Ethics in sustainability; Chapter in book (*Sustainability Methods and Perspectives*, Pressbooks), https://pressbooks.pub/sustainabilitymethods/chapter/ethics-in-sustainability/, Accessed: 2022.

[27] UN Global Compact: Social Sustainability | UN Global Compact; Online resource, 2024, https://unglobalcompact.org/what-is-gc/our-work/social.

[28] E. Thelisson et al.: Toward Responsible AI Use: Considerations for Sustainability Impact Assessment; Preprint (arXiv), December 2023, https://doi.org/10.48550/arxiv.2312.11996.

[29] A. N. P. Kumar et al.: Balancing Progress and Responsibility: A Synthesis of Sustainability Trade-Offs of AI-Based Systems; Conference paper (2024 IEEE 21st International Conference on Software Architecture Companion (ICSA-C), Hyderabad, India, pp. 207–214), 2024, https://doi.org/10.1109/ICSA-C63560.2024.00045.

[30] OECD: How's Life? 2024; Report, 2024, https://doi.org/10.1787/90ba854a-en.

[31] European Commission: Ethics guidelines for trustworthy AI | Shaping Europe's digital future; Online resource, 2019, https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai, Accessed: 2022.

[32] G. Palumbo et al.: Objective metrics for ethical AI: a systematic literature review; Journal article (International Journal of Data Science and Analytics), April 2024, https://doi.org/10.1007/s41060-024-00541-w.

[33] M. Pokholkova et al.: Measuring adherence to AI ethics: A methodology for assessing adherence to ethical principles in the use case of AI-enabled credit scoring application; Journal article (AI and Ethics), April 2024, https://doi.org/10.1007/s43681-024-00468-9.

[34] B. Moldan et al.: How to understand and measure environmental sustainability: Indicators and targets; Journal article (Ecological Indicators, Vol. 17, No. 1, pp. 4–13), June 2012, https://doi.org/10.1016/j.ecolind.2011.04.033.

[35] H. Li et al.: Information synthesis and preliminary case study for life cycle assessment of reflective coatings for cool pavements; Journal article (International Journal of Transportation Science and Technology, Vol. 5, No. 1, pp. 38–46), August 2016, https://doi.org/10.1016/j.ijtst.2016.06.005.

[36] A. Köhler and L. Erdmann: Expected Environmental Impacts of Pervasive Computing; Journal article (Human and Ecological Risk Assessment: An International Journal, Vol. 10, No. 5, pp. 831–852), October 2004, https://doi.org/10.1080/10807030490513856.

[37] A.-L. Ligozat et al.: Unraveling the Hidden Environmental Impacts of AI Solutions for Environment Life Cycle Assessment of AI Solutions; Journal article (Sustainability, Vol. 14, No. 9, pp. 5172), April 2022, https://doi.org/10.3390/su14095172.

[38] Y. E. Wang, G.-Y. Wei, and D. Brooks: Benchmarking TPU, GPU, and CPU Platforms for Deep Learning; Preprint (arXiv), 2019, https://doi.org/10.48550/arXiv.1907.10701.

[39] A. Maria and M. S. Vaija: Life Cycle Analysis of Material Efficiency Strategies for Network Goods; Journal article (International Journal of Automation Technology, Vol. 16, No. 6, pp. 696–703), November 2022, https://doi.org/10.20965/ijat.2022.p0696.

[40] F. Guldbrandsson and P. Bergmark: Opportunities and limitations of using life cycle assessment methodology in the ICT sector; Conference paper (2012 Electronics Goes Green 2012+, Berlin, Germany, pp. 1–6), 2012.

[41] Carnegie Endowment: One of the Biggest Problems in Regulating AI Is Agreeing on a Definition; Online resource, 2022, https://carnegieendowment.org/posts/2022/10/one-of-the-biggest-problems-in-regulating-ai-is-agreeing-on-a-definition?lang=en.

[42] European Parliament: Proposal for a directive on adapting non-contractual civil liability rules to artificial intelligence: Complementary impact assessment; Online resource (Think Tank | European Parliament), 2024, https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2024)762861.

[43] European Union: Regulation (EU) 2024/1689 of the European Parliament and Council; Regulation, 2024.

[44] UNFCCC: The Paris Agreement; Online resource, 2024, https://unfccc.int/process-and-meetings/the-paris-agreement.

[45] M. Z. Chishti et al.: Understanding the Effects of Artificial Intelligence on Energy Transition: The Moderating Role of Paris Agreement; Journal article (Energy Economics, Vol. 131, pp. 107388), March 2024, https://doi.org/10.1016/j.eneco.2024.107388.

[46] Green Energy Report: The Role of AI Technology in the Renewable Energy Sector; Online resource, https://greenenergy.report/articles/the-role-of-ai-technology-in-the-renewable-energy-sector.

[47] S. Ghose: Reducing AI's Climate Impact: Everything You Always Wanted to Know but Were Afraid to Ask; Online article (UC Berkeley Sutardja Center), 2024, https://scet.berkeley.edu/reducing-ais-climate-impact-everything-you-always-wanted-to-know-but-were-afraid-to-ask/.

[48] European Commission: Types of EU law; Online resource, 2023, https://commission.europa.eu/law/law-making-process/types-eu-law_en.

[49] European Commission: Corporate Sustainability Reporting Directive (CSRD); Regulation (Directive 2022/2464 of the European Parliament and of the Council), 2022, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022L2464.

[50] European Union: Energy Efficiency Directive; Regulation (Directive (EU) 2018/2002 of the European Parliament and of the Council), 2018, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32018L2002.

[51] ISO/IEC: ISO/IEC 30134 Series: Data Centre Key Performance Indicators; Standards covering energy efficiency metrics for data centers, including PUE (Power Usage Effectiveness), 2016, https://www.iso.org/standard/71699.html.

[52] J. Xu: Opening the 'black box' of algorithms: Regulation of algorithms in China; Journal article (Communication Research and Practice, Vol. 10, No. 3, pp. 288–296), June 2024, https://doi.org/10.1080/22041451.2024.2346415.

[53] R. Verdecchia et al.: A Systematic Review of Green AI; Preprint (arXiv), January 2023, https://doi.org/10.48550/arxiv.2301.11047.

[54] C.-J. Wu et al.: Sustainable AI: Environmental Implications, Challenges and Opportunities; Preprint (arXiv), January 2022, https://arxiv.org/abs/2111.00364v2.

[55] Energy Institute and Various Sources: Fossil fuel consumption per capita [dataset]; Population based on various sources (2023) – with major processing by Our World in Data; Statistical Review of World Energy, 2024, https://www.energyinstitute.org/statistical-review.

[56] F. Duarte: How Many AI Companies Are There? (2023), Exploding Topics, Jul. 10, 2023, https://explodingtopics.com/blog/number-ai-companies.

[57] C. Clemm, L. Stobbe, K. Wimalawarne, and J. Druschke: Towards Green AI: Current status and future research; arXiv (Cornell University), May 2024, https://doi.org/10.48550/arxiv.2407.10237.

Chapter 7

A QUIC Look at Internet Economics

Ambros Eberhard
Supervised by: Thomas Grübl

As of now TCP is the dominant transport protocol used in the Internet and it was introduced in 1974. Since then, our hardware, software, infrastructure and needs have changed significantly. The first implementation of TCP could not have anticipated the state of our Internet today, therefore there are some pitfalls in TCP when used in today's Internet architecture. Most of these problems can be mitigated by adapting and optimizing the protocol, which has been done over the last 50 years, however this comes at a cost that the protocol will get more and more complex and becomes harder to adapt. QUIC is a rather new transport layer protocol, which aims to solve some of these problems. Specifically, it aims to improve connection-oriented performance. To achieve this, the QUIC protocol tries to reduce the number of round-trips. Furthermore, QUIC is built in a way that it is extensible, so it can be adapted to future needs. However, even if the idea behind QUIC sounds promising, it may introduce new problems that were not existent in TCP. Furthermore, the implications of QUIC on other areas such as economics, privacy, security and resources needs to be evaluated, to be able to conclude whether the idea of QUIC can actually bring significant overall improvements.

Contents

7.1 Introduction

The Internet is a global network composed of computers to transfer information. The basic idea of globally interconnected computers was introduced in 1962. From this idea the Advanced Research Projects Agency Network (ARPANET) was built in 1969, and while it later became the Internet, the original thought was that many independent networks would emerge. Only later did the concept of open architecture networking come along. Open architecture networking is the idea of having independent networks that can provide an interface to interconnect with other networks. To accomplish this interconnection, the Transmission Control Protocol (TCP) was invented [1]. For a long time TCP was the undisputed transport protocol. In 1989 TCP was the dominant transport protocol, accounting for about 80% of packets [2]. Over time the TCP protocol has been adapted many times to improve certain aspects, such as performance, reliability or security [3]. Because of all these improvements and the hardship of rolling out a new protocol TCP is still the de facto standard transport protocol. User Datagram Protocol (UDP) is another transport protocol, which in contrast to TCP is connectionless. Since both connection-oriented and connectionless protocols can be useful, hardware and software in networks have been developed to support TCP and UDP. However the usage of the protocols is not equally distributed. In April 2020, in the MAWI dataset [4], TCP accounted for approximately 90% of Internet traffic while UDP was used only by about 8.5%. Solely 1.5 percentage points of this traffic is sent directly over UDP and the rest runs over QUIC [5].

QUIC is a rather new transport layer protocol, which was developed in 2012 by Google. QUIC aims to replace TCP-over-TLS (Transport Layer Security). The promise is lower latency, better usage of network resources, more extensibility and other benefits [6]. The main goal is to reduce the inevitable drawbacks that arise from problems that could not be anticipated 55 years ago when TCP was developed [7]. The main differences of QUIC and TCP are that QUIC uses HTTP/3 instead of HTTP/2 and it is built on top of UDP making it a transport protocol on top of a transport protocol.

In 2021 QUIC was standardized by the Internet Engineering Task Force (IETF) in the Request for Comments (RFC) 9000 [8]. This laid the basis for the adoption of the QUIC protocol outside of Google. When RFC 9000 was published in May 2021 QUIC accounted for about 12% of global traffic [9] (measurements from Cloudflare radar [10]). Only two years later in April 2023 this number increased to 25% [7] (measurements from Swisscom network monitor and w3techs [11]). Today, it is already at 30% according to Cloudflare [10]. The reason for this fast growth are big tech companies which push the adoption e.g. Google, Meta or Microsoft. Although 30% might sound like a lot it is important to note, that it is used mostly for specific areas such as e.g. Youtube and not everywhere [7]. This leads to the question whether QUIC can fully replace TCP or rather take over specific functions. One problem in answering these questions is the different grades of maturity. While QUIC is mature enough to be implemented in a usable manner in different scenarios, it still has the disadvantage that it did not go through the tedious optimization process which TCP has. Therefore, even if QUICs design would be better suited for a task it might still yield worse results than TCP [12]. While in theory QUIC can provide benefits it is not always clear in practice where these benefits apply. Furthermore the question whom QUIC benefits the most is open [13].

7.2 Background

This section is concerned, with exploring how QUIC was developed, as well as looking at the performance differences between QUIC and TCP. Additionally, a quick look at use cases of QUIC is done in the end.

7.2.1 Development of QUIC

QUIC, originally an acronym for Quick UDP Internet Connections, was first developed in 2012 by Google. Later when the IETF standardization came around the acronym vanished but the name stayed. The development was centered around the following key ideas: Improved performance for HTTPS, fast deployment and good extensibility.
The improvement for HTTPS is needed, since the usage of HTTPS is widespread and the current solution with TCP-over-TLS requires multiple round-trips. This is where QUIC shines. It allows within a single round-trip to have a transport and cryptographic handshake. Furthermore, the client can cache information about the server, to allow for a 0-round-trip time (RTT), where the first packet holds the handshake and encrypted request data, essentially allowing to get data from the server with the first packet sent [14]. In figure 7.1 the 0-RTT functionality of QUIC is illustrated in comparison to TCP + TLS. Note how TCP needs at least one round-trip to establish a connection. TLS also needs two round-trips (in case of initial connection setup) or only one (in case of repeat connection) to establish the security handshake. Only then is the connection ready to request data. QUIC on the other hand can shorten the connection establishment phase and request data within the first packet, in case of a repeat connection.

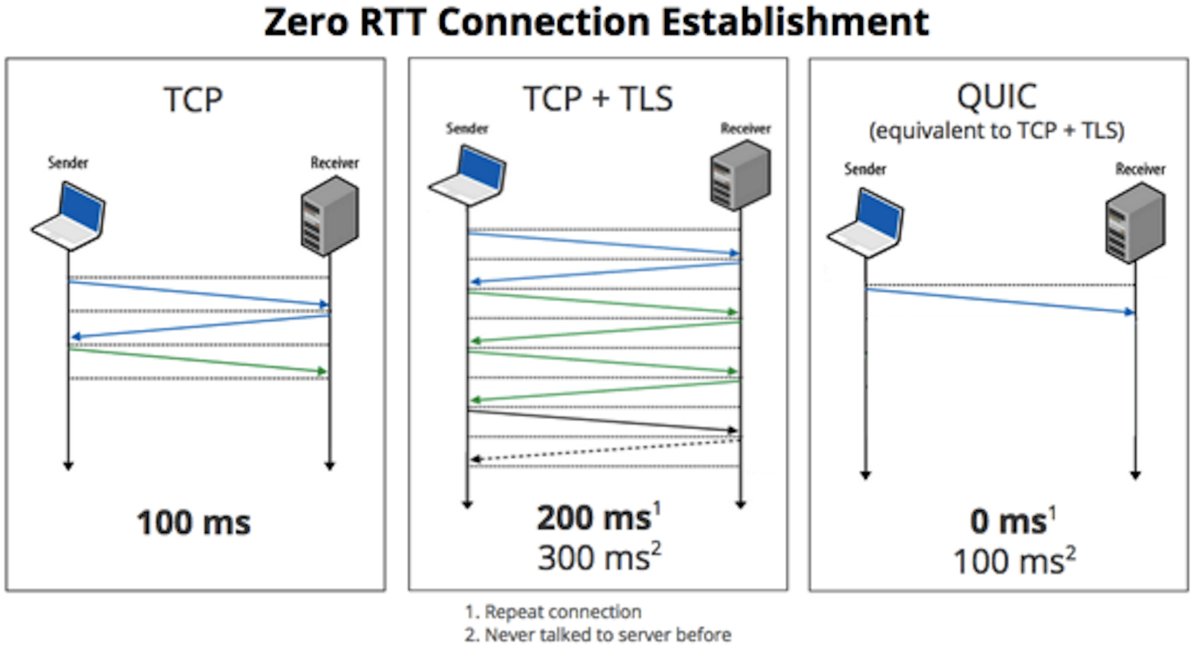


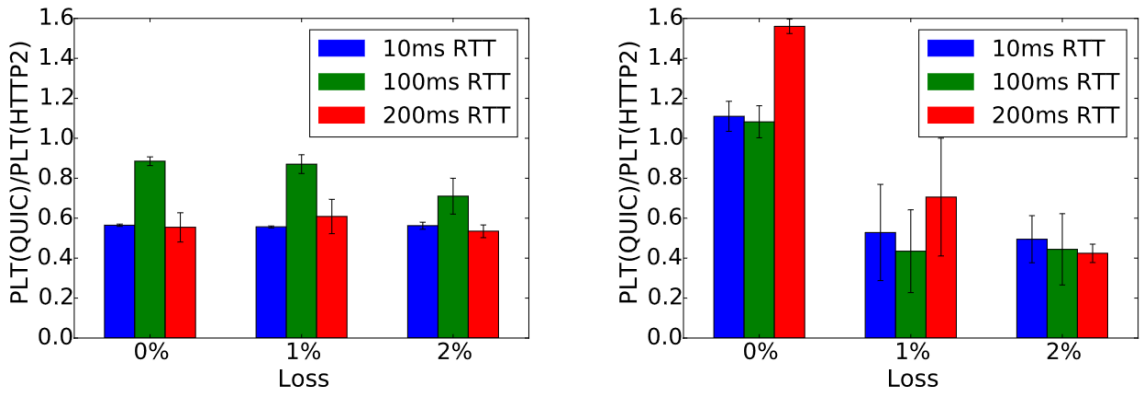
Figure 7.1: Connection establishment of TCP, TCP+TLS and QUIC [15].

Better extensibility and faster deployment are key elements for two reasons. First, it was learned from TCP that transport protocols need to be adapted in the future to meet new requirements. This means the protocol has to constantly adapt, but as seen in TCP changes can lead to protocol ossification which can make future updates to the protocol almost impossible. Second, no updated source code is doing anything unless it is rolled out. TCP is hit hard by this problem, because it is implemented in the kernel space and the middle boxes are programmed for TCP. Due to this design, whenever a change to TCP is made, people have to update their operating system and all the middle boxes have to

be updated as well. QUIC addresses these two problems, by implementing the protocol in user space for faster deployment and embracing an extensible design. While QUIC was first implemented in a monolithic approach, the IETF standard switched to a modular approach, benefiting extensibility. Additionally, the hardware (middle boxes, routers, etc.) need to be able to manage QUIC. To bypass the need to update the software on middles boxes, rather than QUIC implementing its own solution how to manage packets, QUIC runs on top of UDP, allowing global deployment, without changing the firmware/software used by said devices [14].

7.2.2 Performance Aspects

Several aspects slow down TCP and QUIC attempts to improve them. Among them are the handshake delay and the head-of-line (HOL) blocking problem [16]. By reducing the number of round-trips and allowing multiplexed streams in a QUIC connection, QUIC should in theory allow for better performance. For this discussion it is important to know that TCP has existed for 50 years and is widely adopted, which is why it has been optimized a lot, while QUIC is still at its beginning. Therefore, even if better suited, QUIC might still perform worse because of missing optimization. Additionally, there are multiple different implementations of QUIC, which can yield different results [12].



(a) 1 Mbps, 10KB, 64 Objects (b) 100Mbps, 10KB, 64 Objects

Figure 7.2: Page load time of QUIC compared to TCP for different loss and RTT scenarios [16].

QUIC was found to have some applications in which it performs better than TCP. Specifically in high-loss and high-delay scenarios QUIC outperforms TCP in page load time (PLT) as can be seen in figure 7.2 (Note how with just 1 or 2% loss QUIC has a better PLT in the right diagram (b)) [13; 16]. QUIC also has the upper hand on page load time if there is a low bandwidth, as can be seen in the figure 7.2 in the left diagram (a) compared to the right diagram (b) [16]. Furthermore in 4G mobile networks [13]. Lastly QUIC also tends to have better throughput for small files [7]. TCP however performs better in terms of throughput and CPU usage if enough bandwidth is available [17]. It may seem like QUIC is overall a better choice, but it also comes at a cost of 2-3 times the CPU usage of TCP, however again this factor will most likely sink as QUIC will be optimized [7].

7.2.3 Usage of QUIC

QUIC was first developed to reduce browser latency. However, many more use cases came apparent with the standardization of QUIC [18], which is why some companies adopt QUIC internally. In 2017, QUIC was only used internally by Google for its own services

(Gmail, Youtube, Search), in addition, Chrome was the only browser that supported QUIC [13]. In February 2021, Chrome was still the only browser supporting QUIC by default. While other browsers already had support for QUIC it had to be manually enabled. While at this point in time multiple big tech companies such as Google, Microsoft, Facebook and Apple already have adopted QUIC in their servers, still only 5% of websites used QUIC [19]. In June 2021, some companies already relied heavily on QUIC, such as Facebook, which handles 75% of their internal traffic over QUIC. Today all major web browsers support QUIC by default and between 18% to 40% of their requests use QUIC [20]. Yet only 8.5% of websites use QUIC [11], indicating that the websites which get visited more often tend to implement QUIC. Mainly QUIC is used by HTTP/3 and by companies which gain benefits from it.
QUIC is also interesting for mobile networks, because of the use of a connection ID. A connection ID is a 1 to 20 byte number uniquely identifying the connection between two parties. With this ID when either party changes their IP/Port the other side can recognize this by checking the connection ID, which stays the same [7]. This is an advantage over TCP, since with TCP the connection has to be reinitiated every time the 5-tuple changes.

7.2.4 Conclusion

While QUIC can provide substantial benefits in several scenarios [13], for some points alternate measures can be taken to gain the same results. For example better performance can also be achieved by having replicated servers, higher bandwidth or more computing power. Therefore based on the scenario QUIC can make sense or not. However QUICs functionality to be easily extensible and easy deployable is hard to replace. Of course in theory one could rewrite the whole TCP code to make it easier to extend it, however the problem to deploy it would remain. Therefore QUICs approach being implemented in the user space and use UDP as an underlying protocol, is a huge benefit especially for a rather new protocol that needs frequent changes. Further it allows developers to easier adapt the protocol to the individual needs.

7.3 Economic Aspects

In this section the main economic aspects of QUIC will be analyzed. Namely, the questions of which resources are needed to run QUIC and if they are provided. Furthermore, a comparison between costs and benefits of adopting QUIC is drawn.

7.3.1 Resource requirements

Every piece of software has minimum requirements it needs in order to run. For QUIC this includes a minimal network path size. The network path used by QUIC must be able to hold at least 1200 bytes in a single UDP datagram [8]. Furthermore, QUIC needs more CPU power than TCP, since it has not been optimized like TCP and it runs in the user space on top of UDP. Because of this, many system calls are needed, making it computationally expensive [21].

7.3.2 Adoption Costs

Over time the Internet became a huge construct, therefore exchanging an essential part such as the transport protocol (TCP), can easily incur additional costs. However, since this was also known by the developers of QUIC, they created QUIC in a way that it is easy to roll out. Specifically for the required network path size, today's standard for IPv6 and

modern IPv4 already covers the minimum requirements most of the time [8]. The higher CPU usage will incur some additional costs in electricity and possibly in procurement of additional hardware if the currently available is insufficient. Middle boxes do not have to be adapted, since QUIC runs on UDP [14].
About 60% to 70% of the cost for software is caused by maintenance [22]. Therefore, the leading cost factor for QUIC is the development (maintenance/adaptation). Since QUIC is a transport layer protocol and an essential part of the Internet, a breaking change would be fatal. Therefore, the development of the protocol needs to be done with care. In addition, changes must be tested rigorously before deploying [14]. However, these costs can vary based on the initial situation of the company (e.g., fault tolerance, cyber risk, open/closed source implementations).

7.3.3 Benefits

While switching from TCP to QUIC does not directly generate monetary value it can reduce costs or increasing income. These options depend very much on the specific use case. While QUIC tends to use more CPU power than TCP [7; 21], there are certain applications where QUIC can lower CPU usage. Specifically it was shown, that a broker for a Message Queuing Telemetry Transport (MQTT) implementation could achieve lower CPU usage, after publishers were restarted [23]. On the other hand higher incomes can be achieved by adopting QUIC and improving the product quality. This can potentially lead to higher customer satisfaction, which in turn leads to the customers being more likely to spend more money [24]. For this it needs to be analyzed for the specific use case whether QUIC can increase customer satisfaction, as this is not always the case [25].

7.3.4 Conclusion

It is difficult to make general statements about costs and benefits of adopting QUIC, since it depends on the use case. There exists research on when QUIC is useful [13; 16; 17] and some research on user experience when QUIC is adopted [13; 25]. Since the findings are based on different situations it is hard to generalize, but one could produce a mapping from general sample situations to most important factors, based on which the benefits of adopting QUIC can be more easily estimated.
Furthermore, the question if QUIC should indefinitely run over UDP is justified. While it benefits extensibility and easy deployment [14], it also makes the protocol less efficient and induces more operational costs. In the short term it absolutely makes sense to have this, to be able to easily deploy QUIC into production and receive feedback. However, in the long run it might be economically more beneficial to generate an adapted version of UDP for QUIC or even have its own transport mechanism in the kernel. To take this decision, such a mechanism has to be implemented in a minimal form. Then it has to be estimated how far it can be optimized and compared to TCP. The main goal for this implementation would not be to optimize the transport compared to UDP, but minimize the overhead for the system calls.

7.4 Privacy & Security

This part is concerned about the implications for privacy and security when adopting QUIC. It will showcase existing problems and potential countermeasures. Additionally, a comparison to TLS-over-TCP is made.

7.4.1 Privacy

Privacy is an ever growing topic, especially today where user data is collected en mass and processed to build profiles. Therefore privacy concerns need to be addressed in QUIC. QUIC encrypts the whole payload and most of its headers. Only specific QUIC headers, the UDP headers and headers from the lower layers are left unencrypted. While this adds privacy for users by securing the messages, it also gives more privacy to attackers, which might have harmful side effects [7]. QUIC further makes it hard to track users across different connections, since a new Connection Id is issued for each connection [26]. One downside of QUIC is that it can be used to detect which browser is used and which version. This can be done by matching the different QUIC parameters of the initial hello packet against samples of different browsers and versions. However some browsers and versions have the same fingerprint, not allowing to distinguish between them [26]. As of now QUIC can be even more prone to fingerprinting attacks compared to TCP, but QUIC's design allows for countermeasures to be implemented in the future [7]. One big part of why QUIC is useful is because of its connection migration capabilities. However, this connection migration also allows endpoints to track the location of a user [7], which can be used to generate a profile of a person.

7.4.2 Security

Research on the security of QUIC is lagging behind the research on performance, since performance was the main focus in the beginning [27]. Yet, some studies about the security of QUIC have been conducted. These studies [27; 28] have found several potential vulnerabilities which could be exploited. Some are even specifically mentioned in the RFC [8; 29]. These vulnerabilities either stem from QUIC itself or are present in TLS which is used by QUIC. Further, since QUIC runs on top of UDP, problems there can also be used to attack connections over QUIC.
Two big problems with QUIC regarding security are that most of the QUIC data, including metadata, is encrypted and that security measures trade off performance. The problem with the encryption does not lie within QUIC, but rather with how the middle boxes handle traffic. Middle boxes inspect the traffic and analyze header information to take security measures. These security measures can include blocking potential malware and detecting data leaks. Additionally, the encryption of the metadata renders current stateful firewalls useless [7]. There are possible mechanisms to solve this problem, such as using machine learning to analyze traffic patterns and finding abnormalities. Another option would be to use privacy-preserving DPI, where specific tokens are incorporated into the data, allowing the middle boxes to gain information without decrypting. However, no suitable implementation for this exists yet [7].

7.4.3 Comparison to TCP

Not a lot of research exists comparing the security of TLS-over-TCP and QUIC. One study found, that QUIC is about as hard to attack as TCP-TLS [28]. Also, to increase accessibility most sites that feature QUIC can use TCP-TLS which is the fallback if QUIC fails. Therefore, until QUIC is used exclusively for a website, an attacker can use both QUIC and TCP-TLS, basically increasing the possible attack surface [14].

7.4.4 Conclusion

Today not enough research exists, to conclude that one or the other protocol gives more security guarantees. However, once a solution for QUIC is established that replaces the

packet inspection it might surpass TCP. Also currently security measures can slow down QUIC a lot, possibly even resulting in losing the performance advantages of 0-RTT [28]. Maybe in the future when QUIC is more optimized this does not pose a problem anymore, but until then the trade-off between performance and security has to be carefully considered in each use case.

7.5 Company Perspective

This section considers how QUIC can be used from the viewpoint of companies. It will also compare positive and negative aspects of adopting the protocol.

7.5.1 Usage fields in companies

If we take a look at where QUIC performs well, use cases for companies can be derived. Since QUIC performs better in "poor" networks (high data loss, high delay, low bandwidth) [13; 16], it can be useful for new companies that want to keep costs for initial networks low or for international companies which need data transfer between global locations with high delay. However, since TCP has better throughput than QUIC when high bandwidth and a stable network is available, big companies that have sufficient infrastructure might benefit more from TCP. Maybe it would be possible to reduce the bandwidth of the network and use QUIC, to gain the same benefits but at a reduced hardware cost, but to conclude this, further research is needed.
Furthermore, there might be other use cases, which can be beneficial for companies, where QUIC performs better such as transferring small files [7]. Lastly, companies which develop for mobile can benefit from QUIC, since it has better performance on 4G networks [13] and allows for connection migration [7].

7.5.2 Positive Aspects of QUIC for Companies

The adoption of QUIC can benefit companies positively by reducing latency times, therefore reducing waiting times and increasing productivity. However, these benefits can be big or marginal depending on size and use case of the company and are therefore subject to each company's individual possibilities and restrictions. One possible advantage of QUIC is the monopolization of data gathering, which is only on the endpoint and not on the middle boxes - in contrast to TCP [13].

7.5.3 Negative aspects of QUIC for Companies

As seen in Section 7.3.2, adopting QUIC comes with substantial cost. Restrictions and suboptimal decisions of a company might lead to additional costs or a worse solution than TCP (e.g., using an open source implementation of QUIC, when really a personal specific implementation is needed). This poses a risk especially for smaller/newer companies. On top of that, if a company does still have parts of their network running over another protocol than QUIC, the internal IT team at the company has an increased workload, or has to adopt a new subgroup for QUIC specifically.

7.5.4 Conclusion

As mentioned before companies need to have use cases for which QUIC is useful to be adopted. Furthermore, the company needs to be able to provide the resources and expertise to adopt the protocol. Because of TCP still being the standard and not everyone using QUIC, it is usually used as an additional resource. This is probably the reason

why, as of now, mainly big tech firms have adopted QUIC, since they have the resources. However, in the future QUIC might become the standard and it is more lucrative for new businesses to only implement QUIC.

7.6 Individual User Perspective

This section will have a look at the adoption of QUIC from the user perspective. It will highlight use cases, benefits and drawbacks of the adoption.

7.6.1 Usage for Users

While users can make out the difference in speed when using QUIC compared to TCP [25], they do not always feel increased satisfaction. Yet there are scenarios, where QUIC can have substantial benefits such as in mobile networks [13], or video streaming (e.g. QUIC can reduce YouTube rebuffering rates by 15% to 18%) [14] which can possibly lead to increased satisfaction, however this must be further researched.
Another use case is for security and privacy. As discussed in Section 7.4.2, QUIC does not necessarily add security, but it increases privacy for the users.

7.6.2 Positive Aspects of QUIC for Users

As discussed in Section 7.2.2, QUIC can enhance performance in networks with low bandwidth. This can be useful in areas where low bandwidth is available or for private Internet of Things (IoT) devices, which should only use a minimum of resources. Additionally, users get the benefit of added privacy.

7.6.3 Negative aspects of QUIC for Users

Since QUIC is growing rapidly it might soon be used solely for some resources, forcing users to migrate to QUIC even if they do not want to. Further, QUIC monopolizes the data of the user at the endpoint, potentially forming a data source the user does not want. Lastly, the QUIC protocol also brings security risks with it which might lead to breaches.

7.7 Conclusions

As we have seen in Section 7.2, the real power of QUIC lies in its extensibility. This is key for a transport protocol, since we cannot know for certain how the usage or infrastructure of computer networks will be in 20 - 30 years. Therefore, big networks, which have different use cases such as the Internet should try to adopt QUIC. Whereas individual companies need to figure out where adopting QUIC makes sense. As of now, completely replacing TCP in a company would be unwise, since we saw in Section 7.2.2 that TCP performs better over high-bandwidth networks and uses less CPU power as of now. Furthermore, for existing companies to adopt QUIC comes with a significant investment as seen in Section 7.3, why it may not make sense to switch protocols.
While QUIC is rather new it already has great possibilities of application. However, it still has to be optimized, so that it can surpass TCP. During this process, yet unknown flaws will come to light. These flaws will either allow developers to adapt QUIC to handle them or it shows design problems, which contribute to generating a new even better protocol. This is why QUIC needs to be applied widely so enough stakeholders are interested in optimizing it. Adopting QUIC in the Internet would be a great way to accomplish this. Regardless of whether QUIC is widely adopted, we either gain the

benefits of its adaptability and easier deployment or valuable insights by understanding the challenges that prevented its adoption.

Bibliography

[1] B. M. Leiner, V. G. Cerf, D. D. Clark, R. E. Kahn, L. Kleinrock, D. C. Lynch, J. Postel, L. G. Roberts, and S. Wolff, "A brief history of the internet," *ACM SIGCOMM computer communication review*, vol. 39, no. 5, pp. 22–31, 2009.

[2] R. Caceres, *Measurements of wide area internet traffic*. University of California, Berkeley, Computer Science Division, 1989.

[3] A. Medina, M. Allman, and S. Floyd, "Measuring the evolution of transport protocols in the internet," *ACM SIGCOMM Computer Communication Review*, vol. 35, no. 2, pp. 37–52, 2005.

[4] W. Project. Mawi working group traffic archive. [Online]. Available: https://mawi.wide.ad.jp/mawi/

[5] L. Schumann, T. V. Doan, T. Shreedhar, R. Mok, and V. Bajpai, "Impact of evolving protocols and covid-19 on internet traffic shares," *arXiv preprint arXiv:2201.00142*, 2022.

[6] I. TSVAREA, "Quic," 2013, accessed: 2024-11-06. [Online]. Available: https://www.ietf.org/proceedings/88/slides/slides-88-tsvarea-10.pdf

[7] Swisscom, "A quic way to bypass your firewall," Swisscom, Tech. Rep., 2023, accessed: 2024-11-06. [Online]. Available: https://documents.swisscom.com/product/filestore/lib/cbb66c05-4db7-432e-a95b-d6d9523a1c0f/ly2_quic_whitepaper_en_v3.pdf?idxme=pex-search

[8] J. Iyengar and M. Thomson, "QUIC: A UDP-Based Multiplexed and Secure Transport," RFC 9000, May 2021. [Online]. Available: https://www.rfc-editor.org/info/rfc9000

[9] L. Pardue, "Quic version 1 is live on cloudflare," https://blog.cloudflare.com/quic-version-1-is-live-on-cloudflare/, 2021, accessed: 2024-12-19. [Online]. Available: https://blog.cloudflare.com/quic-version-1-is-live-on-cloudflare/

[10] Cloudflare, "Cloudflare radar: Adoption and usage," 2024, accessed: 2024-11-10. [Online]. Available: https://radar.cloudflare.com/adoption-and-usage

[11] W3Techs, "Ce-quic," 2024, accessed: 2024-12-19. [Online]. Available: https://w3techs.com/technologies/details/ce-quic

[12] A. Yu and T. A. Benson, "Dissecting performance of production quic," in *Proceedings of the Web Conference 2021*, 2021, pp. 1157–1168.

[13] S. Cook, B. Mathieu, P. Truong, and I. Hamchaoui, "Quic: Better for what and for whom?" in *2017 IEEE International Conference on Communications (ICC)*. IEEE, 2017, pp. 1–6.

[14] A. Langley, A. Riddoch, A. Wilk, A. Vicente, C. Krasic, D. Zhang, F. Yang, F. Kouranov, I. Swett, J. Iyengar *et al.*, "The quic transport protocol: Design and internet-scale deployment," in *Proceedings of the conference of the ACM special interest group on data communication*, 2017, pp. 183–196.

[15] Philipp Zeder, "QUIC: Neues Protokoll für noch schnellere Websites," 2017, accessed: 2024-11-25. [Online]. Available: https://www.cyon.ch/blog/quic

[16] P. Biswal and O. Gnawali, "Does quic make the web faster?" in *2016 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2016, pp. 1–6.

[17] X. Zhang, S. Jin, Y. He, A. Hassan, Z. M. Mao, F. Qian, and Z.-L. Zhang, "Quic is not quick enough over fast internet," in *Proceedings of the ACM on Web Conference 2024*, 2024, pp. 2713–2722.

[18] A. Technologies. (2021) Http/3 and quic: Past, present, and future. Accessed: 2024-11-18. [Online]. Available: https://www.akamai.com/blog/performance/http3-and-quic-past-present-and-future

[19] F. Networks. (2021) Quic will eat the internet. Accessed: 2024-11-18. [Online]. Available: https://www.f5.com/company/blog/quic-will-eat-the-internet

[20] Cloudflare. (2023) Http/3 usage: One year on. Accessed: 2024-11-18. [Online]. Available: https://blog.cloudflare.com/http3-usage-one-year-on/

[21] T. Shreedhar, R. Panda, S. Podanev, and V. Bajpai, "Evaluating quic performance over web, cloud storage, and video workloads," *IEEE Transactions on Network and Service Management*, vol. 19, no. 2, pp. 1366–1381, 2021.

[22] E. E. Ogheneovo *et al.*, "On the relationship between software complexity and maintenance costs," *Journal of Computer and Communications*, vol. 2, no. 14, p. 1, 2014.

[23] P. Kumar and B. Dezfouli, "Implementation and analysis of quic for mqtt," *Computer Networks*, vol. 150, pp. 28–45, 2019.

[24] C. Homburg, N. Koschate, and W. D. Hoyer, "Do satisfied customers really pay more? a study of the relationship between customer satisfaction and willingness to pay," *Journal of marketing*, vol. 69, no. 2, pp. 84–96, 2005.

[25] J. Rüth, K. Wolsing, K. Wehrle, and O. Hohlfeld, "Perceiving quic: Do users notice or even care?" in *Proceedings of the 15th International Conference on Emerging Networking Experiments And Technologies*, 2019, pp. 144–150.

[26] A. Turner, R. Athapathu, and C. Kharbanda, "Evaluating quic for privacy improvements over its predecessors," 2022.

[27] Y. Joarder and C. Fung, "Exploring quic security and privacy: A comprehensive survey on quic security and privacy vulnerabilities, threats, attacks and future research directions," *IEEE Transactions on Network and Service Management*, 2024.

[28] R. Lychev, S. Jero, A. Boldyreva, and C. Nita-Rotaru, "How secure and quick is quic? provable security and performance analyses," in *2015 IEEE Symposium on Security and Privacy*. IEEE, 2015, pp. 214–231.

[29] S. T. Martin Thomson, "Using TLS to Secure QUIC," Internet Requests for Comments, May 2021. [Online]. Available: https://www.rfc-editor.org/rfc/rfc9001

Chapter 8

5G Techno-Economic Research and its Implications for the Evolution of 6G Wireless Technologies

Alexandru-Mihai Hurjui

As 5G is approaching widespread adoption, researchers are investigating from technical, economic and regulation perspectives the emerging technologies that enable 5G commercial realization. The field of techno-economics, located at the crossroads of these varied perspectives, is an active research area which plays an important role in the design and implementation of a technology. The purpose of this seminar report is to provide an overview of 5G techno-economic literature, and how the current 5G research can offer valuable lessons for the development of next-generation wireless networks.

Contents

Table 8.1: Table of Acronyms

Acronym	**Full Term**
3D-IntCom	Three Dimensional Integrated Communications
3GPP	3rd Generation Partnership Project
5G	Fifth-Generation (cellular network)
6G	Sixth-Generation (cellular network)
AI	Artificial Intelligence
AR	Augmented Reality
ARPU	Average Revenue Per User
CRS	Cognitive Radio System
Capex	Capital Expenditure
D2D	Device-to-Device
DBFA	Dynamic Frequency and Bandwidth Allocation
ELPC	Extremely Low-Power Communications
ETSI	European Telecommunications Standards Institute
FC	Femto Cell (in the context of cellular networks)
FR	Frequency Range
FeMBB	Further Enhanced Mobile Broadband
HetNet	Heterogeneous Network
IEEE	Institute of Electrical and Electronics Engineers
IIoT	Industrial Internet of Things
IMT	International Mobile Telecommunications
IRR	Internal Rate of Return
ITU	International Telecommunication Union
ITU-R	International Telecommunication Union-Radiocommunication Sector
IoT	Internet of Things
KPI	Key Performance Indicator
M2M	Machine-to-Machine
MIMO	Multiple-Input Multiple-Output
MNO	Mobile Network Operator
MVNO	Mobile Virtual Network Operators
NFV	Network Function Virtualization
OTT	Over-the-Top
Opex	Operational Expenditure
QoS	Quality of Service
RF	Radio Frequency
RL	Reinforcement Learning
ROI	Return on Investment
SDN	Software-Defined Networking
SMS	Short Message Service
SurLLC	Secure Ultra-Reliable Low Latency Communications
TCO	Total Cost of Ownership
TEA	Techno-Economic Assessment
THz	Terra-Hertz
UAV	Unmanned Aerial Vehicle
UE	User Equipment
VNF	Virtual Network Function
VR	Virtual Reality
eMBB	Enhanced Mobile Broadband
mMIMO	Massive Multiple-Input Multiple-Output

Acronym	Full Term
mMTC	Massive Machine Type Communication
mmWave	Millimeter wave
uMMTC	Ultra-Massive Machine Type Communication
uRLLC	Ultra-Reliable and Low-Latency Communication

8.1 Introduction

8.1.1 Overview of 5G Techno-Economics

Techno-economics is a field of study that focuses on evaluating a technical system from an economic perspective [2]. Techno-economic evaluations, also known as techno-economic assessments (TEA), combine economic [2], [3], market [4], technological [2], [4], regulation [5] and social [4] aspects to inform the design and deployment of new technologies [2] and their subsequent impacts [4]. Assessments can also use mathematical models, such as Monte Carlo simulations [2] and causal graphs [6], to describe and predict market conditions. This seminar report focuses on the techno-economics of Fifth-Generation (5G) mobile networks and the future trends of cellular network development.

Economic aspects of 5G techno-economics include financial metrics such as capital expenditure (Capex), operational costs (Opex), the Total Cost of Ownership (TCO), Return on Investment (RoI), and Internal Rate of Return (IRR) [2]. Additionally, economic aspects include the impacts of 5G on national and global economies [4].

Relevant aspects from a market point of view include business models for mobile network operators (MNOs), deployment strategies, potential disruptions to the current mobile telecommunication market, and 5G adoption strategies [4].

The technological side of 5G techno-economics refers to the intended use cases of 5G as per the IMT-2020 vision [4], the enabling technologies of 5G [2], its architecture [4], its deployment modes [4], utilized wave frequences [2], and spectral efficiency [2].

From a regulation point of view, techno-economics include environmental protections, spectrum regulation, market power, and national security considerations [5].

Techno-economics can additionally touch on social aspects of using a technology, and in particular the social effect of adopting a new technology. [4] describes examples of social impacts of 5G, including education, smart transportation and innovation. The social side of techno-economics further involves how consumers use the mobile network, and how their usage patterns change over the course of time [4].

8.1.2 Importance for Future Wireless Networks

On a general level, techno-economic evaluation of a technology is important because, after the fundamental research has been done and a standard is developed, the market agents and forces influence the implementation of the standard, the deployment and ultimately the success of the new technology [2]. In particular, 5G techno-economics are important because, by examining all the aforementioned aspects of 5G technologies, researchers, business owners and regulators gain insights into the current state of 5G technology and its ramifications for the future wireless networks [2], [3].

Crucially, techno-economic methods used for 5G can adapted for the future sixth generation (6G) wireless communications. Analyzing the current 5G techno-economic literature reveals its current direction and its weak points, therefore allowing researchers to improve the present-day techno-economic methods and enabling a better assessment of future candidate 6G technologies [2]. The authors of [2], for example, made a survey of the existing 5G techno-economic literature and highlighted key recommendations for the future 6G

research, such more clearly defined technical assumptions, more comprehensive financial metrics and mathematical models, more transparency and more multi-disciplinary cooperation. These recommendations can guide the design and standardization processes of the next generation of cellular technologies, ensuring that 6G can address the current and future challenges of wireless communications.

This insight into analyzing 6G technologies is especially valuable as there is still a significant amount of uncertainty regarding the Next-G wireless communications. The International Telecommunication Union (ITU) and 3rd Generation Partnership Project (3GPP), a union of seven telecommunication standard development organizations [8], are currently working on the 6G standard, which is expected to be finalized between 2026 and 2027 [1]. Researchers are currently exploring the opportunities and challenges that 6G technology is likely to have with respect to use cases, enabling technology and social impact [1], [5]. A clear vision of 6G's direction should be developed which takes into account all stakeholders of the technology [5]. The next-generation telecommunications must additionally support future society's connectivity needs, which require significantly higher data rates, less latency, more mobility and better energy efficiency than what 5G can offer [1]. This can only be achieved with new enabling technology, possible candidates including Edge AI architecture, analysis and planning of network resources using artificial intelligence (AI), and unmanned aerial vehicles (UAVs) which amplify wireless signals [1]. Therefore, techno-economics are an important tool for formulating the vision of 6G and assessing the costs, impacts and feasibility of its enabling technologies.

8.1.3 Objectives of the Seminar

This seminar report aims to offer an overview of the vast field of 5G techno-economic research. Section 2 describes the technological advancements of 5G. Then, we highlight the economic potential and impacts of 5G in Section 3. Business models and deployment strategies are detailed in Section 4. Section 5 describes the future trends of wireless cellular communications. Finally, in Section 6 concludes the paper, presenting the key findings of 5G techno-economic literature and its implications for the future of wireless communications.

8.2 Technological Advancements in 5G and Beyond

The Radiocommunication Sector of the International Telecommunication Union (ITU-R) has identified three main use cases of 5G in its IMT-2020 vision: enhanced mobile broadband (eMBB), considered for high-throughput applications; massive machine type communication (mMTC), meant for a large number of low-power connected devices; and ultra-reliable and low-latency communication (uRLLC), with strong requirements for low latency and reliability [7].

1. eMBB, considered an extension of the highly successful 4G broadband service [4], is developed in the context of rising number of smartphone users and saturation of the current 4G network [2]. eMBB is intended to increase the data rate per cell such that larger numbers of users can consume multimedia content [2]. Other applications include virtual reality (VR), augmented reality (AR), and CCTV [4].

2. mMTC is intended to connect a large amount of low-power and low-rate devices [7], and is intended to be an implementation of the Internet of Things (IoT) [2]. The IMT-2020 vision expects mMTC to support connecting up to 1 million devices per square kilometer [7]. This is especially useful in "verticals", which refers to

industrial sectors such as energy, transport, healthcare, manufacturing, agriculture and construction [2].

3. uRLLC is considered for critical applications which require low latency, high reliability and/or high mobility [2], [4]. Applications considered in this use case include industrial automation, remote surgery, robotics and UAVs [2], [4].

eMBB is currently the only use case that has been successful for MNOs and brought to the market so far [2]. In contrast, uRLLC is, arguably, the most ambitious of the three 5G use cases, but also the one that is farthest from market implementation [2].

8.2.1 Key Innovations Driving 5G Development

These three use cases require new additions or significant improvements to the current telecommunications technology [4]. Techno-economic literature has identified a number of innovations that are expected to enable the realization of the envisioned 5G use cases [2], [4], which can be broadly grouped into three categories: achieving higher network density, higher spectrum bandwidth availability, and greater spectral efficiency [2].

8.2.1.1 Heterogeneous Networks (HetNets) and Ultra-densification

The density of cells can be increased by using a multi-tier cell network [2]. Unlike in a traditional cellular network, HetNets utilize multiple types of cells for different types of communications: macro cells sites provide wide-area coverage, while micro cells are designed for higher data rates in small areas of high demand [2], [4]. For example, the authors of [9] have explored the option of multi-tier networks in the Netherlands, concluding that deployment of small cells only in urban areas is significantly more feasible for MNOs than mass deployment of small cells in the whole country.

8.2.1.2 Software-Defined Networking (SDN) and Network Function Virtualization (NFV)

Previous generations of wireless cellular networks (i.e., 4G and before) were hardware-based and relied on inflexible network architectures [4]. The purpose of SDN and NFV is to virtualize the network functions to run on the cloud, instead of being accomplished by fixed hardware. [4] Software-defined networking (SDN) involves the separation of the user plane (data messages) from the control plane (signaling messages) [4]. This allows more efficient network resource management, as it allows for cloud network control and optimization [4].

Network function virtualization (NFV), a complementary concept to SDN, refers to shifting network functionalities, such as load balancing, from hardware to software components, known as virtual network functions (VNFs) [4]. NFV brings many benefits, including simpler installation, automatic updates and simplyfied network functions [4].

SDN and NFV can bring cost savings of up to 75% compared to traditional hardware-based networks and a 29% increase on ROI [2]. However, the virtualization of the network presents challenges as well, such as security concerns [4] and increases in latency due to the geographical distance to cloud processing facilities [2].

8.2.1.3 Network slicing

An additional network application of network virtualization is network slicing, which refers to partitioning a physical network into multiple virtual, logical networks, with each slice having its own quality of service requirements [2], [4]. Because of the different service requirements, each slice will require its own business model in order to maximize profitability

[4]. This allows transitioning from a "network-as-an-infrastructure" to a "network-as-a-service" business model [4].

8.2.1.4 Millimeter wave (mmWave)

The necessary higher data rates of 5G, in the context of eMBB, can be achieved via usage of increased radio frequency (RF) spectrum resources [4]. Two frequency ranges have been defined by 3GPP for usage in 5G networks: Frequency Range 1 (FR1) contains sub-6Ghz frequencies, and FR2 refers to milimeter wave (e.g., 26 Ghz, 28 Ghz, 60 Ghz) [4]. Milimeter wave has a data rate that is thousands of times higher than sub-6Ghz RF spectrum [4], but has higher propagation losses through walls and atmospheric conditions (e.g., rain) [2]. Therefore, cell ranges are expected to be less than 300 meters, which makes the coverage costs four-five times higher [2].

8.2.1.5 Massive MIMO (mMIMO) and Beamforwarding

Massive multiple-input, multiple-output (mMIMO) is an additional key technology standardized by 3GPP, which aims to maximize spectral efficiency, increasing the data rate [4] and decreasing the per-bit cost of transmission [2]. Massive MIMO consists of base stations equipped with arrays of antenna elements [4], in configurations such as 64-transmit/64-receive (64T64R). Challenges of mMIMO include the higher energy consumption, which raises costs [2]. Researchers recommend a relatively low number of antennas (16T16R or 32T32R) to maintain an energy consumption that remains comercially feasible [2], [10]. However, beamforwarding can increase the capacity and energy efficiency of signals, alleviating some of the energy costs [4].

8.2.2 The Role of AI, IoT and Edge Computing in 5G

Additional significant innovations supporting 5G are AI, the Internet of Things (IoT) and edge computing. AI consists of advanced algorithms that can support essential allocation and maintenance tasks on the network [12]. The IoT addresses the increasing needs for connectivity, and has the potential to play an essential role in industry verticals [2]. Lastly, edge computing is an enabler for low-latency applications of 5G and can be valuable in implementing the uRLLC use case of 5G [11].

8.2.2.1 Artificial Intelligence in 5G networks

As latency and data rates requirements become increasingly stringent, more advanced algorithms must be used for frequency and bandwidth configurations, load balancing, signal relaying and data analysis. Cayamela and Lim [12] highlight how three major categories of AI algorithms can be used for networking tasks.

The first category, supervised learning, is commonly used for predicting future data based on a given training data set [12]. Applications of supervised learning include optimizing the capacity of 5G small cells, which are subject to high amounts of unpredictable interference patterns [12]. The authors of [12] describe how the use of learning-based dynamic frequency and bandwidth allocation (DBFA) prediction models can significantly increase the capacity of these cells and therefore aid MNOs in coping with increasingly higher connectivity demands.

Unsupervised learning, the second class of AI algorithms, is used for detecting groups of related data (often using the *K-means clustering* procedure) [12]. This class can be used for identifying *anomalies*, which are unusual traffic demands for a particular time and location [12]. Anomaly detection can be used for determining locations or regions requir-

ing special attention from MNOs (e.g., additional resource allocation or fault tolerance measures) [12].

Reinforcement learning (RL), which involves an *agent* acting on and reacting to the environment according to *rewards* and *penalties*, is the third major category of AI [12]. The purpose of the agent in RL is to learn a *policy*, which dictates what action it should take at every situation such that the total reward is maximized [12]. RL algorithms find use in femto cells (FCs) in HetNets, where they autonomously adjust their parameters according to the radio environment in order to satisfy the quality of service (QoS) requirements and minimize the intra/inter-tier interference [12], [13].

It is likely that the intersection of AI and 5G technologies will have a significant impact on the future of wireless networks, as the stringent latency and bandwidth requirements of a reliable 5G system require new tools and algorithms [12]. With these AI tools, the future networks can become predictive and AI can play a valuable role in satisfying the increased user needs [12].

8.2.2.2 Internet of Things (IoT) and 5G

An important application of the mMTC use case represents the Internet of Things (IoT) [2], which envisions a large number, potentially millions, of simultaneously connected, interoperable devices [4]. This marks a difference from the approach of traditional cellular networks, where much of the wireless communication takes place between the base station and the device [4]. Enabling technologies of IoT include device-to-device (D2D) and machine-to-machine (M2M) communication innovations [4]. Examples of IoT include, on the consumer side, smart connected homes, and virtual interactions with user environment such as offices, homes and cities [4]. On the industrial side, IoT is relevant for verticals, in which case IoT bears the name Industrial Internet of Things (IIoT) [2]. IIoT is particularly relevant for businesses where mobility (e.g., within a production facility) is necessary, and unlicensed spectrum options (such as Wi-Fi) do not perform up to the expectations [2]. Examples of IIoT applications include waste management, traffic management, water supply management, environmental monitoring and smart agriculture [4].

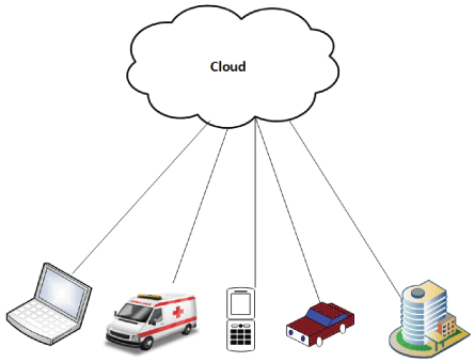
8.2.2.3 Edge Computing in 5G

In traditional cloud computing, the user equipment (UE), which has generally limited processing power and memory, offloads computation and storage to centralized data centers and cloud servers [11]. This centralized model presents problems for 5G, which is foreseen to support low-latency and high-throughput applications [11]. Edge computing is a novel paradigm that aims to satisfy these latency requirements [11]. Edge computing consists of *edge servers* in *mini clouds* that extend the cloud capabilities to perform computationally-intensive tasks and store data close to the user equipment (UE) [11]. Therefore, edge servers have the capabilities of a cloud, but on a smaller scale and located closer to UE than remote data centers [11]. Figure 8.1 (sourced from [11]) illustrates the traditional cloud computing and the edge computing paradigms.

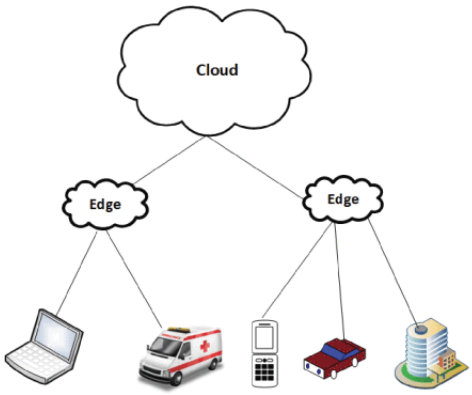
Edge computing supports many applications of 5G, including entertainment and multimedia applications (video streaming and TV), virtual and augmented reality, uRLLC, IoT, emergency response signals and autonomous vehicles [11].

8.3 Economic Impacts of 5G Deployment

The economic impacts of 5G can be expressed with numerous metrics. For this section, we will focus on financial metrics: capital expenditure (Capex), which refers to fixed initial costs (e.g., equipment costs); operational expenditure (Opex), which encompasses



(a) Cloud computing.



(b) Edge computing.

Figure 8.1: Cloud computing and edge computing models [11]

variable costs such as electricity and maintenance; Total Cost of Ownership (TCO), which contains both Opex and Capex; and the Return on Investment (RoI) [2]. Other useful metrics include the average revenue per user (ARPU), which can be used to assess the profitability of different market segments [14]. We will also address the business potential of 5G, and the economic disparities between regions, which impacts 5G adoption [4].

8.3.1 Cost Structures and Investment Requirements

As the authors of [2] show, both Capex and Opex metrics are crucial in techno-economic evaluations of 5G. The Capex of MNOs consists of fixed costs, including civil works, antennas, RF front end and base band as well as labor and transport [14]. The Opex involves site rentals, electricity, licensing, technical maintenance and transport [14].

Distinguishing between market segments is another important aspect of techno-economic analyses. One possible method of segmenting the market is according to the adoption speed of the consumers: when a new service enters the market, the *early adopters* are the first users, and the ones who are the least price-sensitive, and therefore have the highest ARPU [14]. When the service becomes more established, the *mainstream users*, which are more price sensitive, are starting to adopt it [14]. The *laggards* are the least intensive and the most price-sensitive users [14]. Business users are assumed to have an ARPU between the early adopter and the mainstream category of consumers [14].

An additional important factor is the setting in which a 5G network is deployed: additional challenges arise in deploying in a rural environment compared to an urban location, such as lower population density [2]. The economic development of the country considered plays a role, as well: developing markets such as Africa and Latin America present unreliable power supplies, inaccurate demographic data and prohibitive spectrum pricing [4].

These differences in the deployment setting may impose the use of different technologies to ensure a viable level of costs for network operators [2].

The new shift in 5G cost structures to "Anything-as-a-Service" (e.g., "Network-as-a-Service") highlights the importance of accurately estimating the Opex of a network deployment [2]. "Network-as-a-Service" refers to a network operator that opts to rent equipment from a third-party company rather than build their own infrastructure [2]. This business model will become common in the context of network slicing and virtualization, where an MNO does not necessarily own the hardware that they are using [2], [4]. The consequence of this shift is a migration of costs from Capex to Opex, which emphasizes the need for more techno-economic evaluations to include the Opex, considering that Opex analysis is occasionally omitted in scholarly articles [2].

8.3.2 Economic Benefits and ROI for Operators

Researchers have additionally determined the size of the potential 5G markets, and the business stakeholders that will benefit from the 5G transition [4]. By 2026, it is estimated that telecom operators will have an addressable 5G-enabled revenue of $619 billion from ten key industries: healthcare, manufacturing, energy, public safety, agriculture, retail, financial services, transport, media and entertainment and automotive (see Figure 8.2) [4]. By 2030, the expected investment value of 5G across these industries will amount to $700 billion [4]. However, the market opportunities are unequally distributed: Western Europe, North America and North-East Asia offer much higher investment opportunities than other regions, see Figure 8.3 [4]. Expected business stakeholders are not only primary carriers (MNOs), but also secondary carriers, such as mobile virtual network operators (MVNOs), as well as private micro-operators, content providers, app developers, vertical markets and network equipment vendors [4], see Figure 8.4.

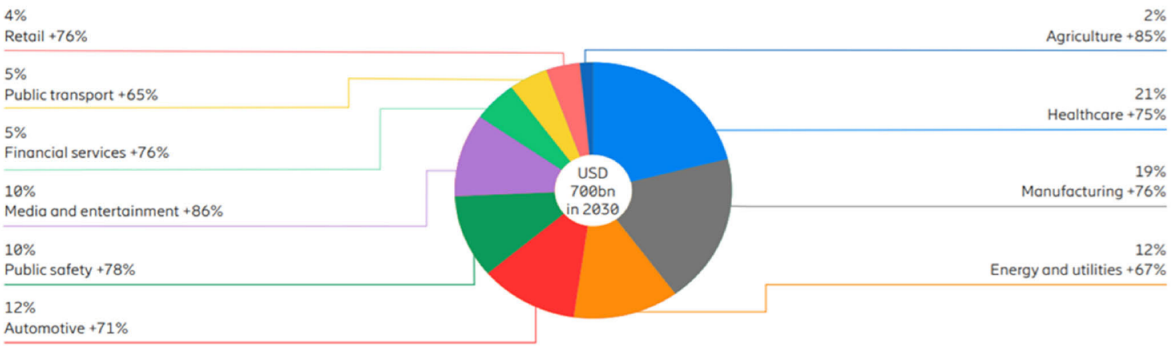


Figure 8.2: 5G-enabled revenue across 10 industries [4]

Techno-economic analysis has revealed the positive implications of 5G for stakeholders [14]. Deployment of eMBB in central London, a highly populated urban area, has been estimated to have a RoI of 29% [14]. Indeed, in the years following the publication of [14], eMBB has seen widespread success among MNOs [2]. The profitability of 5G fluctuates, however, according to the assumed ARPU of each market segment and the expected traffic in the area, as a higher traffic than expected can be detrimental to MNOs [14]. The authors of [14] additionally showed that network sharing has as an almost universal positive effect on revenue levels and RoI.

8.3.3 Societal and Global Economic Implications of 5G

The mobile industry has stated its goal for 5G to "intelligently connect everyone and everything to a better future" [4]. 5G is expected to have an enormous effect on the global

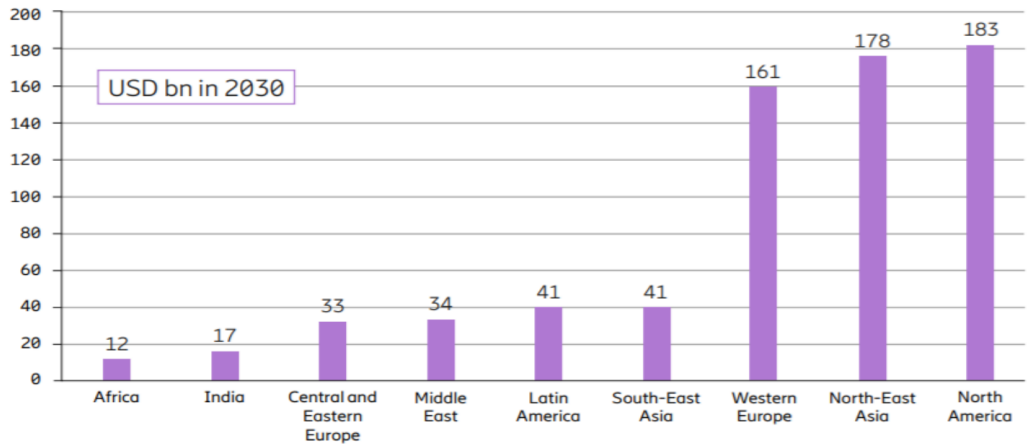


Figure 8.3: Addressable revenue of 5G technology by region [4]

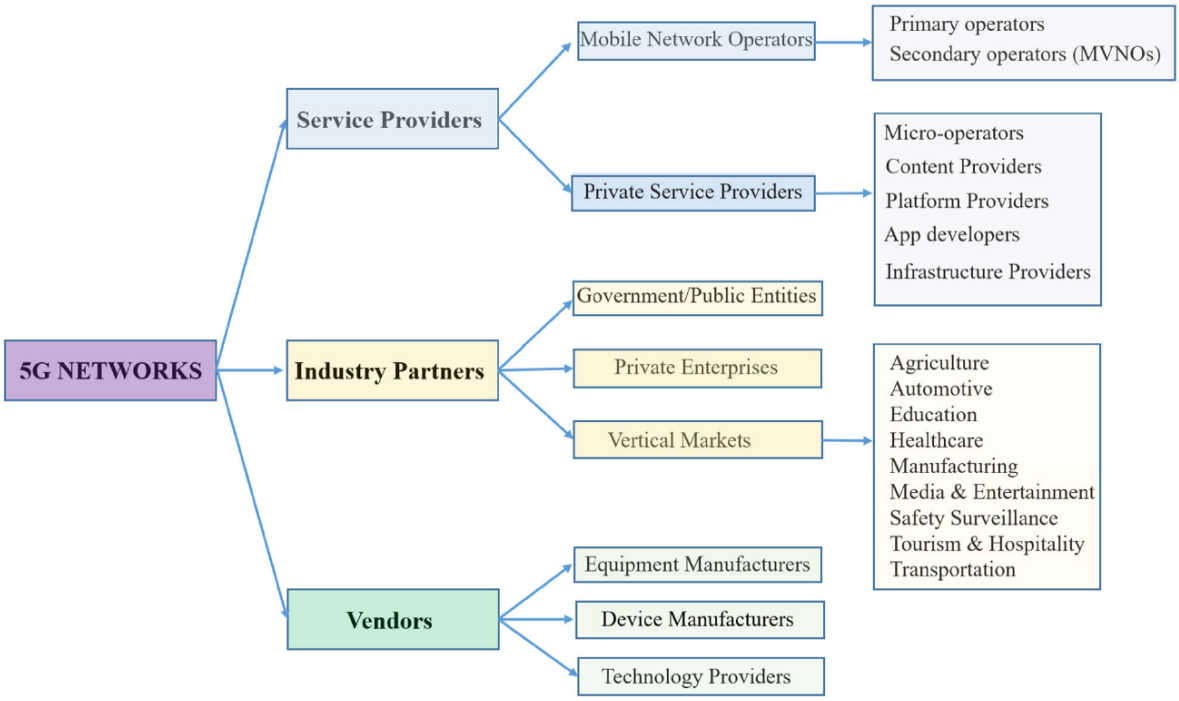


Figure 8.4: Key business stakeholders of 5G [4]

economy, as economists at the World Bank find a strong relationship between broadband penetration and economic growth [4]. It is expected that, between 2020 and 2034, 5G technology will produce economic benefits of $22.2 trillion in the global GDP and bring additional $588 billion in worldwide tax revenue [4]. Another study cited in [4] estimates that 5G will enable $12 trillion of global economic activities in 2035.

Despite the COVID-19 pandemic, the deployment of 5G has continued to grow without interruption [4]. The current coverage of the world population is 35%, while in Western Europe at least 79%, which is in line with predictions from market leaders such as Ericsson [15]. The GSMA predicts that the Asia-Pacific and Sub-Saharan Africa regions will experience the largest growth in new mobile subscribers between 2020 and 2025, although adoption rates will still remain low, at 5% and 3%, respectively [4].

Network operators have the potential to spur the digitization and growth of established and new smart industries through 5G and beyond-5G technologies [4]. The realization of IoT within the mMTC 5G use case will transform vertical industries, such as agriculture, the energy sector and public sector, by enabling the introduction of smart cities,

smart grids and smart irrigation [4]. Considering the rising importance of 5G, tertiary education institutions are intensifying their efforts of teaching the necessary technical and entrepreneurial skills of 5G technology [4]. Examples include the Obuda University in Hungary, which has already introduced 5G technology in their curriculum at both Bachelor and Master levels [4].

8.4 Deployment Strategies and Business Models

Despite its ubiquitous use, the term "business model" lacks a consistent definition and is instead interpreted according to the context of its use [4]. In the context of 5G, a business model can be defined as "[t]he rationale of how a mobile network operator can create, deliver and capture value within the 5G networks' ecosystem by using interconnected elements such as customer relationships, value propositions (services), technology design, financial aspects and infrastructure management" [4]. Clearly, a business model plays a crucial role for an MNO and any company in general, since a business model can be considered a method of implementing a company-wide strategy [4]. Four business pillars must be considered when formulating a business model: the product ("What?"), the customer ("Who?"), the infrastructure management ("How?") and the financial aspect ("How much?") [4].

Until the introduction of 3G and 4G technologies, MNOs were mostly providers of voice communication and short message services (SMS) [4]. The current business models of MNOs additionally consist of mobile broadband, renting their hardware to secondary operators, and entering other markets by acquisition or partnership ventures [4]. However, the disruptive 5G technology and consumer pattern changes bring challenges to these current models, and operators will be required to adapt their business model in order to survive [4].

One of the biggest challenges to the current status is *over-the-top* (OTT) services, which are applications such as social media (e.g., Facebook, Instagram, Twitter), video-on-demand platforms (e.g., YouTube) and messaging applications (e.g., WhatsApp, Telegram) [4]. These services are accessible on the Internet by utilizing the MNOs' infrastructure, without having to own, rent or operate that infrastructure [4]. This is a problem for MNOs, because OTT services represent a threat to traditional voice calls and SMS, which leads to lower revenues for network operators [4].

8.4.1 Market Models for 5G Spectrum Management

The spectrum is considered in many countries a public asset which is controlled by the state [6]. Traditionally, regulatory agencies are the only entity with the authority of assigning radio spectrum to network operators, which takes place in a static manner [6]. This approach results in a spectrum assignment that is almost completely allocated, but underutilized [6]. As the number of devices requiring wireless connectivity increases, the issue of spectrum scarcity and under-utilization becomes more prevalent [6]. Therefore, researchers have suggested the introduction of a secondary spectrum market, where network operators themselves can buy, sell and lease spectrum, leading to a more efficient and flexible allocation of spectrum resources [6], [16].

There are multiple aspects to consider when analyzing secondary spectrum markets. Multiple models for spectrum markets have been proposed, including an auction-based model and commodity trading model [6]. Additionally, spectrum lease durations must strike a challenging balance: shorter leasing durations lead to more efficient spectrum allocation, while longer durations attract more operators and thus foster competition [6].

Novel technologies can also be valuable for establishing a secondary spectrum market [6]. Dynamic spectrum access technologies, such as cognitive radio systems (CRS), allow transmitter devices to change their used frequency bands in an autonomous and real-time manner, in order to exploit the unused spectrum and thus increase the spectrum allocation efficiency [6], [17].

8.4.2 Partnerships, Regulation, and Monetization Strategies

8.4.2.1 Regulatory Components

Regulation is expected to continue playing an important role in the 5G market, as it can foster or inhibit the deployment of 5G [4]. Key regulatory components include:

- Spectrum management: Policy makers should make sure that enough spectrum is available in the appropriate frequency bands with appropriate license conditions [4]. A key issue of spectrum management is the risk of spectrum fragmentation, which can cause interoperability problems [4];
- Network access regulation: Network slicing and "network-as-a-service" models need to feature fair prices to all stakeholders [4];
- Coverage and quality of service (QoS): Regulators should ensure that operators invest the necessary capital to meet the users' QoS requirements [4];
- Network security and privacy: Policy makers should define the rules for data ownership and the what defines data exploitation [4].

8.4.2.2 5G Business Models

Researchers have identified multiple available business models utilizing 5G technology effectively:

1. Vertical partnership business models: network operators can enter partnerships with stakeholders in various verticals in order to capture value through a collaborative value system [4]. An example is smart agriculture, where MNOs partner with Ministries of Agriculture and commercial farmers to introduce 5G-based IoT to connect wireless sensors and irrigation systems [4]. Similarly, smart electricity grids can help Ministries of Energy and power utility companies effectively monitor and forecast energy demand using 5G networks [4].
2. OTT service providers: MNOs can opt to compete with traditional OTT providers by building their own platform (e.g., own video streaming service), or by partnering with a third-party OTT for developing the platform [4].
3. Smart infrastructure: Private 5G network operators can sell network solutions to factories for enabling machine-to-machine communications, to sea ports for automated cranes, to railway companies for better rail signaling and to mine companies for easier deployment of sensors which monitor the work environment [4].

8.4.3 Regional and Global Deployment Case Studies

[14] shows that deployment of 5G in central London, a dense urban population of a Western country, brings significant profit to operators, with an expected 29% ROI when investing into 5G eMBB. Additionally, network sharing schemes have an almost universally positive effect on the revenue stream [14].

Developing countries are mostly still an untapped market for 5G network operators, considering the low rate of adoption in Sub-Saharan Africa and Asia-Pacific [4]. 5G deployment is more challenging for MNOs in these markets compared to highly-developed countries, considering the lack of adequate demographic data, unreliable electric grids and low fiber penetration [4]. Indeed, Europe enjoys a relatively high fiber coverage rate of 62% [3], which makes the technology a viable option for backhauling [4]. In contrast, a study case in Nigeria highlights the power supply shortage in the country and the importance of electricity savings when considering backhaul options [18]. Considering the absence of fiber penetration in Nigeria, the authors have identified microwave, V-band, E-band and self-backhauling as viable backhaul options for the country of 200 million inhabitants [18].

8.5 Future trends: 6G Evolution

Despite the existing uncertainty around 5G investment options and business models, researchers and industry have already started designing use cases, performance requirements and technical specifications for the next generation (6G) of cellular wireless networks [2]. Indeed, [3] highlights five growing axes of 6G literature, exploring the vision, use cases, Key Performance Indicators (KPIs), business models and TEAs of 6G networks. As the authors of [2] state, this is "a familiar, if messy, process". Nevertheless, the existing 5G literature is sufficient to allow researchers to analyze the trends of 5G evolution and already formulate plans for the 6G technology [2].

One must be aware of the fact that, since there is no 3GPP release for a 6G standard yet [1], the current visions of 6G are competing and partially overlapping, and its requirements still ambiguous [5]. This suggests the need of a unified vision framework for 6G, which takes into account all stakeholders (technology, regulation and business) to ensure a sustainable approach to 6G development [5].

8.5.1 Lessons from 5G Techno-Economic Research for 6G

Based on a comprehensive survey of 5G techno-economic literature, experts have formulated five key recommendations for evaluating 6G candidate technologies [2]:

- Quality of service assumptions used in TEAs (e.g., traffic demand, interference, spectral efficiency etc.) must be much more clearly stated in order to boost research clarity. The assumptions made in a TEA heavily influence its results and therefore its conclusions. The literature survey has revealed the worrying trend of assumptions being communicated in places that researchers may miss.

- Useful financial metrics must to included in the TEA. Some research papers did not include the Opex of a 5G deployment, which leads to inaccurate results and inefficient decision making. The authors recommend including the TCO in TEAs.

- TEAs must include sensitivity analyses to quantify 6G model uncertainties. The purpose of a sensitivity analysis is to ensure that the assumptions and uncertainties of a model are fully portrayed, and to account for different deployment contexts.

- Researchers must openly share their 6G model data and code. This is to ensure transparency and high standards of scientific inquiry, as it allows other researchers to evaluate, validate and inspect the key contributions of a paper.

- Greater multi-disciplinary cooperation is needed in the research, standardization and techno-economic assessment processes. The authors argue that 5G has received too little techno-economic attention during the early R&D and standardization process, a mistake which should be avoided for 6G.

8.5.2 Emerging Requirements and Vision for 6G

There is a large number of requirements and visions of 6G described in the research literature [5]. In general, 6G is imagined to have an even more profound effect on society and business than 5G [5]. The vision of 6G is to revolutionize mobile networking and support the needs of a future, data-driven society [1]. 6G is expected to support the convergence of physical and digital worlds by means of e.g., holographic representations and digital twins [1], [5]. The infrastructure is expected to be "smart", by developing AI-based algorithms that allow the network to allocate resources, manage and maintain itself autonomously [1]. The next-generation networks are anticipated to integrate with satellites as well, allowing more satellite-based services such as navigation, weather forecasting and earth imaging [1]. Lastly, 6G is anticipated to finally connect every person to the Internet, since there are still 2.9 billion people on the planet with no Internet access [5].

6G is anticipated to be more efficient, reliable, scalable and energy-efficient than 5G [1], with data throughput of up to 1 Tbps, latency of less than 1ms and availability of more than 99.99% [1], [3].

8.5.3 Potential Technology Enablers and Future Standards

8.5.3.1 Technology Enablers of 6G

There is a wide number of anticipated 6G enablers [1], including:

- UAV-assisted wireless communications: UAVs can be used as aerial base stations or relay nodes for supporting terrestrial and aerial 3D wireless communication [1]. This brings significant benefits compared to fixed terrestrial base stations, and increases the network resilience and capacity by enabling a flexible node topology [1].

- Satellite communications: Satellites are an alternative to UAV-based networks for assisting terrestrial communications [1]. The 3GPP is currently working on standardizing satellite transmission for this purpose [1].

- Terra-Hertz (THz) wave: The current mmWave (30-300 GHz) employed in 5G cannot support the massive data rate requirements of 6G [1]. Therefore, researchers have started to explore the options of higher frequency ranges of up to 6 THz [1].

- Artificial intelligence: AI can be used for analyzing the colossal amounts of data produced by IoT devices, helping to improve the system performance [1]. AI algorithms can additionally be used for signal processing, network design, resource allocation and self-maintaining infrastructure [1].

8.5.3.2 Future Standardization

Use cases of 6G that could be standardized include Further enhanced Mobile Broadband (FeMBB), Secure ultra-reliable Low Latency Communications (SurLLC), ultra-massive Machine Type Communication (umMTC), Extremely Low-Power Communications (ELPC), and Three Dimensional Integrated Communications (3D-IntCom) [1]. Organizations involved in the development of 6G include ITU, 3GPP, the European Telecommunications Standards Institute (ETSI), the Institute of Electrical and Electronics Engineers (IEEE), governments such as Japan, and private companies such as Huawei, Samsung Electronics, LG, Sony and many more [1].

A 6G standard is currently being developed by ITU and 3GPP, and is expected to be finalized between 2026 and 2027, while network trials are forecast for the years 2028 to 2030 [1]. Figure 8.5 shows a likely 6G evolution timeline for the 2020 decade [1].

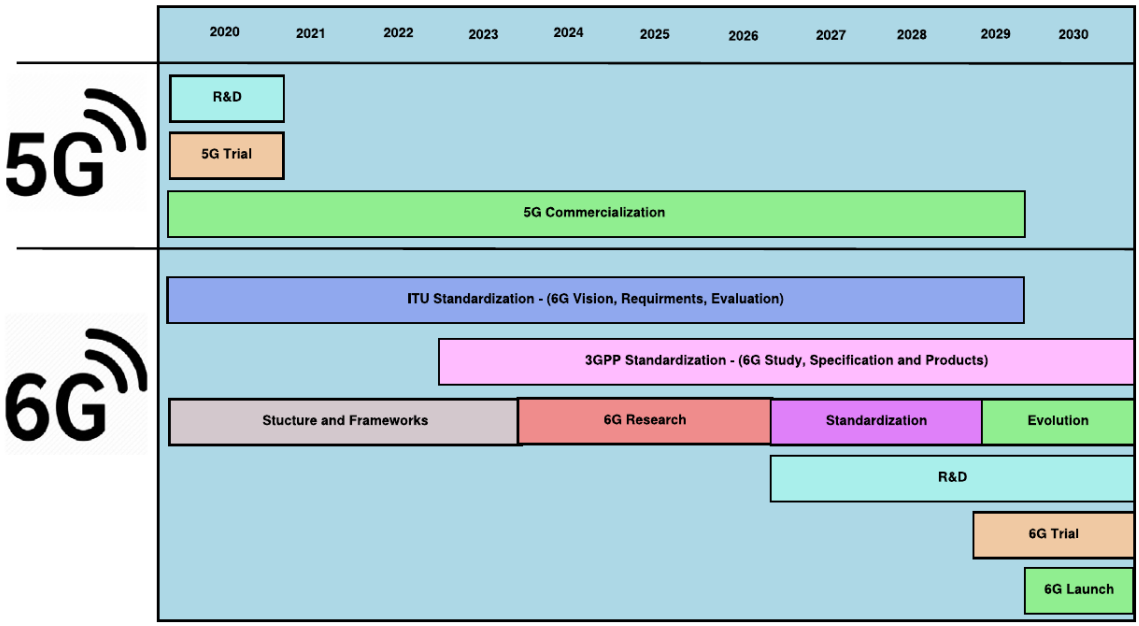


Figure 8.5: Expected 6G evolution timeline [1]

8.5.4 Anticipated Societal and Industrial Impacts

Understandably, a technological leap such as 6G will bring significant societal changes. Digitalization and software will be ubiquitous, and the scarcity of developer capacity will be exacerbated [5]. A more scalable and power-efficient 5G can connect rural and poorer regions, finally addressing the stark rich-poor and urban-rural gaps seen in 5G deployment [1], [2], [4], [18]. This connection brings great social and economic development, as these markets are still mostly untapped [2], [4].

The advent of always-connected, ”zero-lag” devices and systems will transform numerous verticals by automating crucial activities or making them more reliable [1]. Additionally, 6G will create new market ecosystems surrounding these industries, enabling more cooperation between stakeholders such as equipment providers, content providers, regulators, resource brokers and security providers [1].

8.6 Conclusion

This seminar report has presented 5G cellular networks from the main perspectives of technology, economics and business, while touching upon the social and regulatory aspects. Based on this overview of 5G techno-economics, we shall present our key findings and the prospects of future wireless networks.

8.6.1 Summary of Key Findings

5G is still a relatively new technology, and we are far from the end of its lifecycle [2]. Out of the three main use cases formulated in the IMT-2020 vision, only one (eMBB) has found successful deployment so far [2], and mostly just in developed economies, which are more urbanized, where customers have a higher willingness-to-pay and where the environment presents fewer business and technical obstacles [4]. The positive effects of eMBB 5G deployment in urban areas have been proved by numerous TEAs [2], such as [14], while less developed areas, such as rural areas, have received less attention from techno-economic literature [2]. This is problematic, considering that rural regions encompass 3

billion people [2] and there is certainly market potential in developing countries such as Nigeria [18].

Considering this relatively limited deployment of a single use case, 5G has achieved only a fraction of its enormous economic, social and business potential. Many applications such as AR, VR, and IIoT still need more research to reach market viability. Indeed, the 3GPP has standardized uRLLC in later releases than eMBB [2], and the techno-economic literature has given less attention to mMTC and especially to uRLLC compared to eMBB [2]. Granted, 5G presents new technical challenges compared to previous cellular generations. Stringent latency requirements encourage research to explore new network architectures such as edge computing [11]. The increasingly softwarized networking functions present great flexibility and performance benefits [4], [12], but also introduce security challenges [4]. In the context of increasingly numerous connected devices, researchers explore changes to the current state-dominated spectrum model in pursuit of more efficient spectrum utilization [6]. Similarly, researchers have suggested new business models for MNOs to overcome the disruptive effect of 5G on the wireless communication market, but there are still open research questions on the details of many proposed business models [4].

Despite this uncertainty of 5G, however, experts have already started to formulate requirements, use cases and technical specifications of 6G cellular networks [2]. 6G is an increasingly broad field in literature, as five axes of research have been identified in literature: visions of 6G, use cases, KPIs, business models and TEAs [3]. Papers such as [1] describe the numerous breakthrough technologies that could enable 6G, the applications of 6G networks, and how the society and businesses will benefit from this technology. However, as there is no standardization of 6G yet, there is no clarity on the included technologies in 6G. Therefore, the current literature on 6G is arguably vague, and many of the envisioned use cases and requirements are partially overlapping and competing [5].

8.6.2 Implications for Future Wireless Technologies

As the authors of [2] expressed it, we are "not close to the end of 5G's lifecycle", but rather "at the end of 5G's beginning". While the 5G literature still presents research gaps, it is sufficiently vast to extract trends and lessons from the research of 5G which will be valuable for analyzing future 6G technologies [2].

Based on the current 5G techno-economic literature, experts have issued five key recommendations to improve the research process of next-generation wireless networks: clear communication of model assumptions, comprehensive financial metrics, modeling uncertainty using sensitivity analysis, openly sharing model data and code, and greater multidisciplinary collaboration in early phases of 6G development [2]. Indeed, development frameworks such as the one presented in [5] can ensure that the early research and R&D processes of 6G produce a coherent vision of a sustainable, human-centric technology which takes into account all stakeholders.

8.6.3 Final Thoughts on 5G-6G Transition

As time progresses, 5G will be fully deployed commercially and eventually become a legacy technology in the 2030s, while the requirements and enabling technologies of 6G will be sufficiently explored to enable network trials and more robust research on next-generation wireless networks [1]. However, researchers can maximize the positive social and economic impacts of next-G networks only if they apply the lessons learned in 5G development. For example, 5G has not addressed the low penetration rate of wireless networks in rural environments, which presents 6G with the opportunity for finally "connecting the unconnected" [2], [5]. While 5G is not yet a fully mature technology, it already features in a

significant amount of literature and certainly has the potential to inform the development of future wireless networks [2].

Bibliography

[1] D. K. R and S. Chavhan: "Shift to 6G: Exploration on trends, vision, requirements, technologies, research, and standardization efforts"; Sustainable Energy Technologies and Assessments, vol. 54, p. 102666, Dec. 2022.

[2] E. J. Oughton and W. Lehr: "Surveying 5G Techno-Economic Research to Inform the Evaluation of 6G Wireless Technologies"; arXiv.org, 06-Jan-2022. [Online]. Available: https://arxiv.org/abs/2201.02272.

[3] D. Kokkinis, N. Ioannou, D. Katsianis, and D. Varoutas: "A 6G Techno-Economic Framework for evaluating the feasibility of the proposed technology enablers and business models"; 2023. [Online]. Available: https://www.econstor.eu/handle/10419/277988.

[4] L. Banda, M. Mzyece and F. Mekuria: "5G Business Models for Mobile Network Operators - A Survey,"; in IEEE Access, vol. 10, pp. 94851-94886, 2022, doi: 10.1109/ACCESS.2022.3205011

[5] P. Ahokangas, M. Matinmikko-Blue and S. Yryola: "Envisioning a Future-Proof Global 6G from Business, Regulation, and Technology Perspectives,"; in IEEE Communications Magazine, vol. 61, no. 2, pp. 72-78, February 2023, doi: 10.1109/MCOM.001.2200310

[6] L. A. Fletscher, A. Zuleta, A. Galvis, D. Quintero, J. F. Botero, and N. Gaviria: "A Techno-Economic Analysis of New Market Models for 5G+ Spectrum Management"; Information, vol. 15, no. 4, p. 197, Apr. 2024.

[7] ITU-R: "IMT Vision - Framework and overall objectives of the future development of IMT for 2020 and beyond"; Recommendation ITU-R M.2083-0, Sept. 2015, available online under: https://www.itu.int/rec/R-REC-M.2083

[8] 3GPP: "Introducing 3GPP"; website retrieved on 2024-11-17, URL: https://www.3gpp.org/about-us/introducing-3gpp

[9] Edward J. Oughton, Zoraida Frias, Sietse van der Gaast, Rudolf van der Berg: "Assessing the capacity, coverage and cost of 5G infrastructure strategies: Analysis of the Netherlands", Telematics and Informatics, Volume 37, 2019, Pages 50-69, ISSN 0736-5853, doi: https://doi.org/10.1016/j.tele.2019.01.003

[10] Gedel, I.A.; Nwulu, N.I." "Low Latency 5G Distributed Wireless Network Architecture: A Techno-Economic Comparison."; Inventions 2021, 6, 11. https://doi.org/10.3390/inventions6010011

[11] N. Hassan, K. -L. A. Yau and C. Wu, "Edge Computing in 5G: A Review," in IEEE Access, vol. 7, pp. 127276-127289, 2019, doi: 10.1109/ACCESS.2019.2938534

[12] M. E. Morocho Cayamcela and W. Lim, "Artificial Intelligence in 5G Technology: A Survey," 2018 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, Korea (South), 2018, pp. 860-865, doi: 10.1109/ICTC.2018.8539642

[13] G. Alnwaimi, S. Vahid and K. Moessner, "Dynamic Heterogeneous Learning Games for Opportunistic Access in LTE-Based Macro/Femtocell Deployments," in IEEE Transactions on Wireless Communications, vol. 14, no. 4, pp. 2294-2308, April 2015, doi: 10.1109/TWC.2014.2384510

[14] Juan Rendon Schneir, Ade Ajibulu, Konstantinos Konstantinou, Julie Bradford, Gerd Zimmermann, Heinz Droste, Rafael Canto: "A business case for 5G mobile broadband in a dense urban area"; Telecommunications Policy, Volume 43, Issue 7, 2019, 101813, ISSN 0308-5961, doi: `https://doi.org/10.1016/j.telpol.2019.03.002`

[15] S. Mishra, A. F. Zanella, O. E. Martínez-Durive, D. Madariaga, C. Ziemlicki and M. Fiore: "Characterizing 5G Adoption and its Impact on Network Traffic and Mobile Service Consumption," IEEE INFOCOM 2024 - IEEE Conference on Computer Communications, Vancouver, BC, Canada, 2024, pp. 1531-1540, doi: 10.1109/INFOCOM52122.2024.10621344

[16] Hee Seok Song, Taewan Kim, Taehan Kim: "The impact of spectrum policies on the secondary spectrum market: A system dynamics approach"; Telecommunications Policy, Volume 41, Issues 5-6, 2017, Pages 460-472, ISSN 0308-5961, doi: `https://doi.org/10.1016/j.telpol.2017.04.004`

[17] Arturo Basaure, Vladimir Marianov, Ricardo Paredes: "Implications of dynamic spectrum management for regulation"; Telecommunications Policy, Volume 39, Issue 7, 2015, Pages 563-579, ISSN 0308-5961, doi: `https://doi.org/10.1016/j.telpol.2014.07.001`

[18] G. Salami, N. Faruk, N. Surajudeen-Bakinde and F. Ngobigha: "Challenges and Trends in 5G Deployment: A Nigerian Case Study"; 2019 2nd International Conference of the IEEE Nigeria Computer Chapter (NigeriaComputConf), Zaria, Nigeria, 2019, pp. 1-6, doi: 10.1109/NigeriaComputConf45974.2019.8949675

Chapter 9

The Economics of Digital Twins in Clinical Scenarios

Linda Weber

The healthcare sector faces significant challenges, including rising costs, inefficiencies, and the limitations of generalized treatment approaches. Digital twins, as system-of-systems, offer a transformative solution through real-time data integration, continuous monitoring, and proactive treatment. By leveraging technologies such as internet of things (IoT) and artificial intelligence (AI) digital twins enable personalized care and efficient resource allocation, addressing both clinical and economic needs. This paper conducts a systematic review of 25 studies, analyzing enabling technologies, application scenarios, and regulatory considerations. Key applications include personalized medicine, surgical simulation, chronic disease management, and drug development. Despite their promise, digital twins face challenges such as high implementation costs, data integration issues, and ethical concerns. However, their potential to revolutionize healthcare through patient-centered, proactive, and economically sustainable solutions highlights their importance in shaping the future of the sector.

Keywords: Digital twins, healthcare, clinical scenarios, proactive treatment, personalized care, resource allocation, artificial intelligence (AI)

Contents

9.1 Introduction

In recent years, the shift towards digital technologies has accelerated significantly. The Covid-19 pandemic has acted as a further catalyst for the digitalisation in many industries. It has forced organizations to make decisions faster and based on real-time data. This digital transformation is a critical part of industry 4.0, also called the fourth industrial revolution, that emphasizes automation, data exchange and smart systems [6]. Furthermore, there have been considerable recent technological advancements. Artificial intelligence (AI), internet of things (IoT), sensor technology and other advancements have helped break down barriers and deliver better technological solutions. At the same time, healthcare costs pose a great challenge for many nations. Costs are rising due to factors like the aging populations and high numbers of chronic diseases, making healthcare one of the most expensive sectors globally. These developments have made it clear that there is a need for new technologies that can better support patients and doctors and optimize the healthcare sector. Enter digital twins: a system-of-systems that has the potential to revolutionize healthcare. The term digital twin is not new but has rather regained attention recently. Its origins date back over 50 years, to the Apollo 13 mission in 1970, where NASA engineers created a virtual replica of the spacecraft to analyze its status in real-time and test solutions to guide the spacecraft safely back to earth. The concept was not called digital twins at that time, but it demonstrated the powers of such a technology. Michael Grieves further developed the concept for product lifecycle management and named it "Mirrored Spaces Model" in 2002 [17].

Since then, digital twins have seen successful implementations in different industries, particularly in manufacturing and aerospace. These early successes have set the foundations for broader and more complex applications in other areas like smart cities, logistics, and healthcare.

Currently, many different definitions of digital twins exist in the literature, but for this work, the definition proposed by [17] will be used:

'A digital twin is a self-adapting, self-regulating, self-monitoring, and self-diagnosing system-of-systems with the following properties: 1.) It is characterized by a symbiotic relationship between a physical entity and its virtual representation. 2.) Its fidelity, rate of synchronization, and choice of enabling technologies are tailored to its envisioned use cases. 3.) It supports services that add operational and business value to the physical entity.' [17, p. 7]

This definition was chosen for this paper for several reasons: it fits many different applications, (not just in healthcare, but across various sectors), it is comprehensive and it will likely remain relevant for the foreseeable future. In the context of healthcare, a digital twin can be a representation of a process such as trauma management, or it can be the representation of a human organ, or even a representation of an entire patient. Applications are discussed in detail in chapter 9.4.

Using the definition from above, the meaning of a digital twin in healthcare is best illustrated with an example. To optimize cardiovascular treatment, a virtual replica of the patient's heart and vascular system, including demographic information, clinical data and continuous sensor readings can be created [20]. The (1) *symbiotic relationship* refers to the ongoing exchange of data between the patient and their digital twin, ensuring that the virtual model mirrors the physical reality. The (2) *fidelity and synchronization* describe how accurate and how frequently the digital twin updates to reflect its physical counterpart. The requirements depend heavily on the use case. For example, in cardiovascular health, where decisions might depend on real-time data like heart rate or ECG signals, the digital twin needs high fidelity and near-constant synchronization to remain effective. On the other hand, in something like a digital twin for structural monitoring of a building, updates might occur every few hours or even daily because changes happen

much more slowly. Technologies for digital twins are discussed in depth in chapter 9.3. And finally, (3) the *value* is realized through improved outcomes for patients and more efficient healthcare systems.

Today's healthcare often relies on periodic check-ups and generalized treatments, which means that interventions may come too late, and treatments aren't always optimized for individual needs. Digital twins could replace this one-size-fits-all approach and enable three key advancements. First, proactive monitoring and prevention: By continuously analyzing the patient's virtual twin and running predictive analyses, clinicians can detect anomalies early and intervene before issues escalate. Secondly, personalized treatment: Each patient is unique, and an optimal treatment journey requires personalized interventions based on factors such as the patient's genetics, lifestyle, and medical history. Such personalized treatment improves outcomes as well as minimizes side effects. And the third benefit is efficient resource allocation: With digital twins, healthcare providers can deliver care efficiently to minimize costs and maximize outcomes.

Previous works have covered many aspects of digital twins in healthcare, such as suggesting new research directions, exploring existing frameworks and proposing improved ones, discussing ethical, societal and regulatory challenges, and of course, presenting in-depth case studies of specific digital twin applications in healthcare. Other works have taken a broader approach and reviewed existing scientific papers on digital twins and their technologies, advantages, and challenges across different sectors. However, there is a lack of comprehensive surveys analyzing the economic implications of digital twins in healthcare. This type of work is essential and valuable because the healthcare sector is currently not sustainable and long-term solutions to optimize care and minimize costs are crucial. By carrying out a systematic literature review and analyzing previous work across multiple dimensions (research type, main technologies, key applications, regulatory and ethical aspects, advantages and challenges), this paper aims to address the current research gap. It also contributes by presenting an identification of open challenges and directions for future research, emphasizing the need for interdisciplinary collaboration and standardization. These contributions aim to provide a comprehensive overview that bridges technological innovation and economic feasibility in the context of digital twins. Therefore, it has the potential to aid decision-makers, shape healthcare policies and foster further innovation and discussion of this research topic.

9.2 Research Methodology

For this seminar paper, a structured literature review of recent studies about the economics of digital twins in clinical scenarios was conducted. The process was as follows: first, Google Scholar was chosen as the search engine. Secondly, significant keywords were identified: the search result must have the words "Digital Twin" in the title, and at least one of the following words "Healthcare Clinical Economics" in the title. This resulted in the following search using boolean operators: "Digital Twin" AND ("Healthcare" OR "Clinical" OR "Economics"). Third, the time frame was set to 2021-now (October 2024). This decision was made for two reasons: 1) to make sure that the papers included newer technological advancements in artificial intelligence and sensor technology and to 2) refine the scope of the work. This search yielded 130 results. As a next step, a categorization of the results was carried out manually. Previous work that is highly relevant for this seminar paper was categorized as List A. Specific examples and case studies were categorized as List B. And papers that were not relevant were categorized as List C. Categorization was done by reading the title (for some specific examples, it was clear they were in List B, even from just the title), if that did not suffice, abstracts and keywords were read, and in some cases, the conclusion. This resulted in 21 papers categorized as List A (see appendix

Source	Research Type and Scope	Main Technologies	Key Application Areas/ Clinical Use Cases	Key Regulatory/ Ethical Aspects	Advantages of Digital Twins in Healthcare	Challenges of Digital Twins in Healthcare	Other
Abd Elaziz et al., 2024	Review article, covering applications, technologies, and future trends	IoT, AI, VR, Big Data, Cloud Computing, Simulations	Healthcare 4.0, Cardiac analysis, Monitoring, Data privacy, Surgical applications	Addresses privacy issues through anonymization, encryption, and access controls	Enhanced diagnostic accuracy, personalized medicine, improved resource allocation	Data privacy concerns, complexity in implementation, lack of standardization	Suggests new research directions for future DT applications
Adibi et al., 2024	Comprehensive review of digital twins in smart environments. Focus on IoT and AI technologies.	IoT, AI, IoMT, Machine Learning, Wearables, Location-based services (LBS)	Remote patient monitoring, telemedicine, smart hospitals, personalized healthcare interventions	Data privacy concerns with real-time data sharing; standardization of data protocols for interoperability	Real-time health monitoring, early detection of health issues, optimized resource allocation, personalized treatments	Complexity in data integration, privacy concerns, lack of standardization, and high computational requirements	Explores a proposed framework for integrating digital twins in smart environments for healthcare optimization
Alazab et al., 2023	Overview and proposal of a new architecture for digital twins in Healthcare 4.0, covering recent advances and challenges	IoT, AI, Blockchain, Edge Computing, Cloud Computing, Machine Learning, Wearables	Personalized healthcare, Telemedicine, Intelligent rehabilitation, Smart diet management	Data privacy, interoperability, standardization	Improved resource optimization, real-time monitoring, enhanced patient outcomes	High computational demands, data privacy, complexity of system integration, lack of standardization	Proposal of a new layered architecture for digital twins in healthcare
Amram et al., 2023	Ethical commentary on the implications of digital twins and clones in healthcare	Digital twins, AI, Data Simulation	Drug development, Personalized medicine, Clinical trials alternatives	Data ownership, privacy rights, consent for downstream use, moral implications of experimentation	Enhanced drug development processes, potential alternatives to double-blind trials	Data ownership disputes, privacy concerns, risk of unethical experiments, complexity of consent	Focus on bioethical challenges; potential AI-driven sentience in digital twins, clones
Armeni et al., 2022	Comprehensive review exploring the potential of digital twins in healthcare, with a focus on precision medicine and clinical trials	IoT, AI, Cloud computing, Big data, Simulation	Precision medicine, Clinical trials design, Hospital management	Privacy concerns, data security, equity issues, accessibility of DTs technology	Personalized treatment, improved clinical trials design, optimized hospital operations	Data bias, cost of implementation, lack of standardized data, challenges in real-world implementation	Suggests the use of digital twins for social equity issues and advocates for further studies to validate their use in clinical trials
Balasubramanyam et al., 2024	Literature review from 2020 to 2023 examining the roles and benefits of digital twins in personalized healthcare	AI, IoT, Machine Learning, Big Data, Cloud Computing	Personalized healthcare, predictive healthcare, virtual clinical trials, diagnostics, and treatment optimization	Privacy concerns, data integration, access control, informed consent, security issues related to real-time data sharing	Personalized medicine, real-time health monitoring, optimization of treatment plans, reduced cost of clinical trials	Data integration complexity, lack of standardization, high implementation costs, concerns over data privacy and security	Provides an in-depth analysis of how digital twins impact predictive healthcare and resource optimization in hospitals
Batty, 2018	Editorial commentary	IoT, AI, GIS, Building Information Models (BIM)	Urban planning, city management, infrastructure maintenance, transportation systems	Lack of regulatory frameworks for urban digital twins, potential surveillance issues, data privacy concerns	Real-time management of urban environments, optimization of infrastructure, enhanced predictive modeling	High complexity in creating real-time models, lack of comprehensive data, difficulty in integrating socio-economic factors with physical infrastructure	Focuses on applying digital twins to cities rather than healthcare; discusses broader technological and philosophical implications of digital twins
Bordukova et al., 2024	Review article discussing how generative AI empowers digital twins in drug discovery and clinical trials	Generative AI, Machine Learning (ML), Neural Networks (NN), GANs, VAE	Drug discovery, Clinical trials, Preclinical testing, Predictive modeling, Organ simulations	Data privacy, regulatory approval, validation of AI-generated data in clinical settings, transparency issues in AI models	Accelerates drug discovery, improves trial accuracy, enhances patient safety, reduces animal testing	Lack of transparency in AI models, limited regulatory guidelines, difficulty in real-world implementation of digital twins in clinical trials	Emphasizes the role of generative AI in creating virtual patients, offers a vision for future integrated AI-driven digital twin systems
Croatti et al., 2020	Research on integrating multi-agent systems (MAS) with digital twins in healthcare, focusing on trauma management	IoT, AI, Multi-Agent Systems (MAS), Simulation, Cloud Computing	Trauma management, real-time monitoring, patient care, pre-hospital care, hospital resource optimization	Data privacy, interoperability, ethical concerns around autonomy in AI systems	Real-time trauma monitoring, improved resource allocation, enhanced coordination between pre-hospital and hospital teams	Data security, complexity in system integration, challenges in real-time data processing	Case study of a digital twin for trauma management; emphasizes the potential for agent-based digital twins to improve healthcare processes
De Maeyer & Markopoulos, 2021	Delphi study	IoT, AI, Wearables, Cloud Computing, Electronic Health Records	Preventive healthcare, Personalized medicine, ICU monitoring, Clinical trials, Population health tracking	Data privacy, autonomy, human-in-the-loop, transparency, ethical guidelines (EU AI ethics)	Enables preventive healthcare, personalized treatment, trial-and-error simulations, enhanced decision-making	Interoperability of systems, data security, patient autonomy, technological immaturity	Focus on expert consensus from a Delphi study in Belgium; highlights preventive healthcare as a key advantage of DTs
Johnson & Saikia, 2024	Review of digital twin applications in healthcare	Wearable sensors, IoT, AI, Machine Learning, Blockchain	Remote monitoring, real-time disease progression tracking, personalized medicine, point-of-care testing	Data privacy, data security, interoperability between devices, ethical use of patient data	Real-time health monitoring, personalized treatment, faster diagnosis, remote healthcare	Sensor limitations, real-time data processing, interoperability, patient privacy, cost of sensors and wearables	Proposes a method for generating holistic human body digital twins, highlights potential applications in personalized treatment and disease monitoring

Figure 9.1: Systematic literature comparison across multiple dimensions. Part I.

A), meaning they are the most relevant. An additional 4 papers were put into List A, as they were suggested in the topic description by the seminar professor. In total, 45 papers were put into List B (specific examples/case studies), see appendix B, and a total of 64 papers were put into List C (discarded), see appendix C.

Source	Research Type and Scope	Main Technologies	Key Application Areas/ Clinical Use Cases	Key Regulatory/ Ethical Aspects	Advantages of Digital Twins in Healthcare	Challenges of Digital Twins in Healthcare	Other
Korovin, 2022	Exploratory research on digital twins in industry, assessing their maturity, applications, and economic impacts	IoT, AI, Cloud Computing, Sensors, Multidisciplinary Modeling	Industrial production, product life cycle management, manufacturing, process optimization	Data security, economic feasibility, potential information overload	Enhanced product quality, optimized business processes, predictive maintenance, reduced costs	High initial cost of implementation, technological immaturity, data integration challenges	Focuses on the impact of digital twins on industrial transformation and economic effects
Kshetri, 2021	Exploratory study on the economic, social, and environmental benefits of digital twins	IoT, AI, Machine Learning, Big Data Analytics, Blockchain	Internal organ simulations, virtual patients, predictive healthcare, in silico clinical trials	Data privacy, regulatory approval for AI-based interventions, data security	Faster regulatory approval for drugs, improved patient safety, reduced animal testing, enhanced process efficiency	High costs of developing high-fidelity models, technical limitations, data accessibility	Explores the economic impact of digital twins across various industries, with a focus on healthcare applications like virtual patients and in silico clinical trials
Kuriakose et al., 2024	Bibliometric analysis of 1,663 publications on digital twins in healthcare from 2012–2024	IoT, AI, Machine Learning, Data Analytics, Wearables	Personalized medicine, chronic disease management, predictive analytics, surgical planning	Data privacy, data interoperability, patient consent, model validation	Enhanced diagnosis, optimized treatment plans, real-time health monitoring, proactive care	Interoperability issues, data privacy concerns, validation of models across diverse populations	Highlights bibliometric trends and research gaps; emphasizes the fast-growing nature of digital twin research since 2018
Marfoglia et al., 2024	Conceptual framework to integrate machine learning models within digital twins for enhanced simulation capabilities in personalized healthcare	Machine Learning (ML), Knowledge Graphs (KGs), AI, IoT, Semantic Technologies	Personalized healthcare, Predictive healthcare, Proactive care, Chronic disease management	Data privacy, security in data integration, transparency of ML models, ethical issues in automation	Enhanced predictive models, personalized treatments, real-time patient data simulation, improved decision-making	Integration complexity, challenges in data standardization, need for transparent ML models, high resource requirements	Proposes a novel framework for combining ML models with digital twins to create more dynamic and responsive systems for personalized healthcare
Meijer et al., 2023	Narrative review focused on the methodological challenges of digital twins in healthcare, highlighting data integration and standardization issues	IoT, AI, Machine Learning, Multi-omics, Medical Imaging, Signal Processing	Personalized medicine, chronic disease management, drug development, treatment strategy optimization	Data privacy, data standardization, model validation, complexity of integration	Enhanced personalized care, real-time monitoring, predictive analytics, reduced need for animal testing	Standardizing multi-source data, high computational demand, issues with accuracy and reliability of models	Comprehensive look at the methodological challenges, with case studies in different healthcare areas (diabetes, cardiovascular, cancer research)
Mihai et al., 2022	Survey on enabling technologies, challenges, trends, and future prospects for digital twins across industries	IoT, AI, 5G/6G, AR/VR, Blockchain, Cloud Computing, Transfer Learning	Personalized medicine, remote monitoring, surgical simulation, predictive healthcare	Data privacy, standardization, interoperability, data ownership, security	Real-time monitoring, predictive analytics, optimized treatment, reduced costs	High costs, lack of data standardization, interoperability issues, lack of clear regulatory guidelines	Focuses on digital twin applications in Industry 4.0, with insights into various sectors, including healthcare
Mohamed et al., 2023	Explores how digital twins can improve healthcare systems engineering, focusing on process optimization and decision-making	IoT, AI, Sensors, Cloud Computing, Edge Computing, Machine Learning, Cyber-Physical Systems (CPS)	Personalized healthcare, resource management, predictive maintenance, process optimization, patient care, risk management	Data privacy, data validation, noise filtering, automation, ROI evaluation, system abstraction, interoperability	Real-time monitoring, predictive analytics, process optimization, cost-effectiveness, enhanced decision-making	Data collection challenges, noise reduction, high cost of implementation, complexity in creating patient-specific digital twins, lack of standardization	The paper proposes a conceptual framework and emphasizes the importance of digital twins for system optimization and future decision-making
Popa et al., 2021	A socio-ethical analysis exploring the potential benefits and risks of digital twins in healthcare	IoT, AI, Big Data, Robotics, Simulation, Cloud Computing	Personalized medicine, clinical trials, diagnostics, surgery planning	Data privacy, patient autonomy, inequality, informed consent, data ownership	Enhanced patient autonomy, better diagnosis, cost reduction, improved clinical trials	Privacy concerns, healthcare inequality, social disruption, algorithmic bias	Highlights both benefits and ethical risks, with recommendations for governance frameworks and interdisciplinary regulation
Sharma et al., 2024	Empirical study focusing on a machine learning-based ECG classifier for healthcare applications. Integrates real-time data processing with IoT and AI in healthcare.	IoT, Artificial Intelligence (AI), Machine Learning (ML), Neural Networks (e.g., LSTM, CNN), Cloud Computing	Cardiac diagnostics and monitoring, early anomaly detection, continuous health tracking, predictive maintenance for patient-specific applications.	Addresses data privacy concerns and emphasizes the need for secure handling of patient data.	Enhanced diagnostic accuracy, real-time monitoring, scalability of health systems, improved patient-physician communication, better use of predictive analytics.	Reliability of ML models, data privacy and security concerns, integration challenges, lack of standardization for large-scale implementations.	Highlights the potential of using neural network-based methods to improve healthcare outcomes. Sets a precedent for integrating AI in continuous health monitoring.

Figure 9.2: Systematic literature comparison across multiple dimensions. Part II.

9.3 Digital Twin Technologies

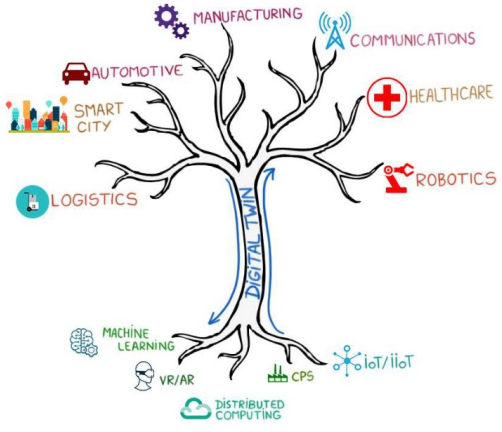


Figure 9.4: Illustration of the digital twin concept as a tree, from [17, p. 1].

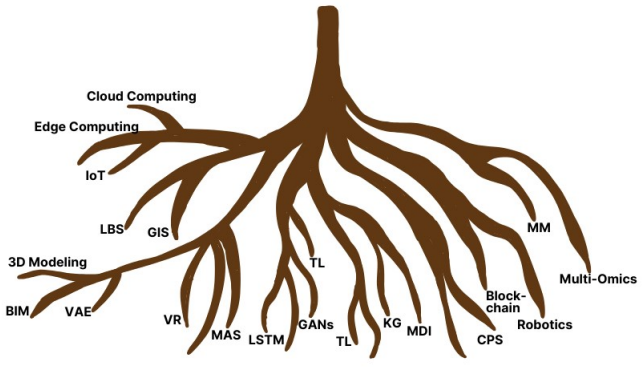


Figure 9.5: Own figure, adapted from [17]. Shows a more detailed view of digital twin technologies, abbreviations see table in C.1.

Source	Research Type and Scope	Main Technologies	Key Application Areas/ Clinical Use Cases	Key Regulatory/ Ethical Aspects	Advantages of Digital Twins in Healthcare	Challenges of Digital Twins in Healthcare	Other
Subasi & Subasi, 2024	Comprehensive review of digital twins in healthcare and biomedicine, focusing on integration with AI, IoT, and blockchain technologies. Covers applications, challenges, and future directions.	IoT, AI, Machine Learning (ML), Blockchain, Big Data Analytics, and Federated Learning.	Personalized medicine, precision therapies, oncology treatment optimization, disease monitoring, and cardiovascular disease management.	Data privacy concerns related to the integration of IoT and AI; ethical issues in patient-specific digital twin development; need for standardized frameworks.	Enables real-time monitoring, personalized treatment, improved diagnostic accuracy, and data-driven healthcare optimization.	High computational requirements, data security risks, interoperability challenges, lack of clinical trial integration, and ethical dilemmas (e.g., discrimination, inequity).	Highlights the potential of hybrid models (mechanistic and data-driven) for combining multiple healthcare data types; discusses applications for cancer and cardiovascular care.
Turab & Jamil, 2023	Comprehensive survey exploring the intersection of digital twins and the metaverse in healthcare, focusing on personalized and precise medicine, and immersive care.	IoT, AI, Blockchain, Virtual Reality (VR), Augmented Reality (AR), Big Data Analytics, Edge Computing, and 5G connectivity.	Personalized medicine, virtual counseling, precision therapies, medical education, drug development, and immersive patient monitoring.	Highlights ethical concerns like informed consent, data ownership, and ensuring fairness; emphasizes interoperability and standards in global adoption.	Facilitates precision and predictive medicine, enhances patient engagement through immersive experiences, and allows for seamless integration of care models.	Technical complexities with metaverse integration, ethical dilemmas in immersive care, high costs of implementation, and interoperability barriers.	Links digital twin applications to metaverse technologies, highlighting transformative potential in education, therapy, and chronic disease management.
Vidovszky et al., 2024	Position paper (meaning not purely research-based) discussing AI-generated digital twins and their role in clinical trials, emphasizing their potential to transform drug development and clinical practice.	AI, Machine Learning (ML), Deep Learning, Prognostic Covariate Adjustment (PROCOVA), and Multimodal Data Integration.	Clinical trial optimization, predictive modeling for treatment outcomes, drug development acceleration, personalized medicine, and rare disease research.	Regulatory support needed to align digital twin usage in clinical trials with current standards; privacy concerns about patient data and model explainability.	Enhances efficiency in drug trials, reduces trial sizes while maintaining statistical power, and supports personalized medicine with tailored predictions.	Challenges include data standardization, handling complexity in biological systems, and developing robust validation frameworks to address regulatory scrutiny.	Advocates for increased adoption through regulatory alignment; links DT use to reducing trial costs and increasing accessibility in rare disease research.
Volkov et al., 2021	Review paper analyzing digital twins, IoT, and mobile medicine platforms in healthcare. Proposes a concept for a unified Smart Healthcare Platform.	IoT, Mobile Medicine, Cloud Computing, Edge Computing, Wearable Devices, and Machine Learning (ML).	Continuous health monitoring, telemedicine, disease prevention, personalized medicine, and mobile health applications.	Discusses risks of data breaches, patient data ownership issues, and challenges with standardizing healthcare data exchange.	Improves patient monitoring and engagement, supports disease prevention, and enables personalized medicine through real-time data insights.	Interoperability challenges, lack of unified platforms for health data exchange, and high costs for integrating multiple technologies across systems.	Proposes a conceptual architecture for a Smart Healthcare Platform, emphasizing modularity and secure data-sharing between patients, providers, and researchers.
Zhang et al., 2024	Comprehensive analysis of digital twin concepts and applications in healthcare, with a focus on their integration into personalized medicine and smart diagnostics.	IoT, AI, Big Data Analytics, 3D Modeling, Cloud Computing, and Blockchain Technologies.	Smart diagnostics, surgical planning, predictive healthcare modeling, and patient-specific treatment planning.	Highlights concerns about informed consent, data security, patient privacy, and the need for global regulatory standards for healthcare digital twins.	Improves accuracy in diagnosis, facilitates personalized treatment plans, enhances patient outcomes, and integrates predictive analytics into routine care.	High initial investment costs, lack of interoperability, regulatory barriers across regions, and challenges in scaling models to diverse patient populations.	Focuses on the potential of blockchain for securing healthcare data; emphasizes the importance of collaboration between technologists and clinicians.

Figure 9.3: Systematic literature comparison across multiple dimensions. Part II.

In their work, Mihai et al. [17] presented the concept of digital twins as a tree visualization, see 9.4. The roots show the enabling technologies of the digital twin systems, like machine learning and internet of things. The branches show some applications of digital twins, such as logistics and healthcare. This visualization is great for helping with understanding and remembering key facts about digital twins. Therefore, it was used and adapted for this paper.

Based on the papers reviewed (see figures 9.1, 9.2, 9.3), the root was extended to include additional technologies, see figure 9.5. They can be categorized into six key areas of technology that form the foundation of digital twins: Data Acquisition and Integration, Simulation and Modeling, Advanced Analytics and Artificial Intelligence, Communication and Connectivity, Security and Privacy, and Hybrid and Emerging Technologies.

9.3.1 Data Acquisition and Integration

Data acquisition and integration involves gathering and combining data from multiple sources. Technologies like the internet of things, enable real-time data collection through wearables, sensors, and medical devices. For example, sensors embedded in wearable devices transmit patient-specific data, such as heart rate or blood pressure, to a digital twin of the individual, enabling healthcare providers to monitor conditions remotely and intervene proactively [21]. Edge computing processes some of this data locally on devices, providing faster insights by reducing the need to send everything to the cloud. On the other hand, cloud computing offers scalability by centralizing data for analysis and storage. Geographic information systems is technology, that collects, analyzes, and visualizes spatial and geographic data. In healthcare, it has various applications, from helping to model the spread of infectious diseases [24], to identifying optimal locations for healthcare facilities [23].

9.3.2 Simulation and Modeling

Simulation and modeling is about creating dynamic and interactive virtual representations. For example, 3D modeling allows experts to build virtual replicas of organs for surgery planning or diagnosis [24]. One application of this is that medical students can use these interactive 3D models to practice surgical techniques, reducing risks in live scenarios [22]. Virtual and augmented reality add another layer, enabling immersive environments for training or enhanced visualization during procedures such as surgeries. Multi-agent systems can simulate how different entities, like hospital staff or patients, interact in real-world scenarios to optimize workflows. Another related technology is building information models (BIM), which can be used design and manage hospital facilities, including layouts, energy systems, and patient flow optimization [24]. Furthermore, variational autoencoders (VAE) are machine learning models that compress and reconstruct data and can help with reconstructing high-fidelity digital representations of organs from MRI or CT scans [26].

9.3.3 Advanced Analytics and Artificial Intelligence

AI and machine learning are at the heart of digital twin intelligence. Tools like convolutional neural networks are able to reliably and accurately analyze medical images, saving time and reducing human errors. Artificial intelligence is also a critical enabler for advancing the integration of wearables and sensor data. AI can process large volumes of sensor data, and enhance data reliability. Advanced approaches, like generative adversarial networks, can even create synthetic data for training models when real-world data is insufficient [11]. Long short-term memory (LSTM) networks are a type of recurrent neural network and excel in assisting with the prediction of patterns in time-series data. In digital twin technologies in healthcare, this is extremely useful for the continuous analysis and monitoring of data such as heart rate and oxygen levels of the patient [21]. However, training these networks on large datasets requires high computational resources and therefore increases costs. Transfer learning, federated learning, prognostic covariate adjustment (PROCOVA), and advanced data integration technologies (knowledge graphs and semantic technologies, multimodal data integration) were also mentioned in some of the papers analyzed, but will not be explained here. For further information, the corresponding papers can be found in figures 9.1, 9.2, 9.3.

9.3.4 Communication and Connectivity

Digital twins rely on seamless and secure communication. 5G and emerging 6G technologies provide the speed and low latency required for real-time data exchange [22]. Cyber-physical systems (CPS) integrate computational systems with physical processes, therefore bridging the gap between the digital and physical worlds and enabling continuous monitoring and adjustments [25].

9.3.5 Security and Privacy

Due to the highly sensitive patient data, the implementation of technologies that ensure privacy and security is key. One such technology is blockchain, because it can address critical challenges such as the need for immutable and auditable records. Blockchain in the context of digital twins in healthcare also has the ability to enhance traceability by linking physical twins with their digital counterparts in a transparent ledger. However, the use of blockchain in this context faces many challenges still, such as the high latency and power consumption [13].

9.3.6 Hybrid and Emerging Technologies

Mobile medicine integrates mobile devices such as smartphones, tablets, and wearables into healthcare workflows and acts as an extension of digital twin systems, facilitating continuous data collection and interaction [22]. It can reduce costs by streamlining communications, reducing the need for in-person visits, and facilitating decision-making. However, the required infrastructure, the security of these systems and the accessibility can be expensive and complex to develop and deploy. Multi-omics is a complex biological analysis approach, which in the context of healthcare, can provide a detailed, multi-dimensional view of a patient's health. Applications of multi-omics include drug discovery and early detection of genetic predispositions [26].

9.4 Digital Twin Application Scenarios

There are many different digital twin applications in healthcare, some of which have already been in use for several years, and others that are just a concept or prototypes for now. In this section, an overview of digital twin applications, based on the literature reviewed is presented. The findings are summarized in table 9.1, and explained in more detail in the following subsections. The overview is not exhaustive and is expected to further expand in the future. For now, it shows the most frequently mentioned applications of digital twins in healthcare.

Table 9.1: Overview of digital twin application areas in digital health.

Category	Subcategory	Examples
1. Healthcare	a. Personalized Medicine	[20], [26]
	b. Predictive Healthcare	[20], [26]
	c. Surgical Simulation	[1]
	d. Diagnostics	[20]
	e. Chronic Disease Management	[20]
2. Systems Engineering	a. Resource Management	[18], [9]
	b. Process Optimization	[18], [9]
3. Monitoring and Remote Care	a. Remote Monitoring	[2]
	b. Telemedicine	[2]
4. Drug and Devices Development	a. Preclinical Testing	[8]
	b. Organ Simulations	[8]
	c. Medical Devices	[26]

9.4.1 Healthcare

One example of digital twins in healthcare is that of cardiovascular health [20], where the digital twin is created for the patient's heart and vascular system. This example illustrates, that such a digital twin application often covers multiple subcategories to provide the best possible treatment. By integrating the patient's specific demographic information, clinical data and continuous sensor readings, the digital twin enables personalized care (1.a.) and therefore optimizes treatment. Furthermore, the digital twin is used to run predictive analyses (1.b.) and help with making an (updated) diagnosis (1.d.), and if needed, assist manage the patient's chronic cardiovascular disease (1.e.). Digital twins offer a way to carry out risk-free surgery simulations [1] for especially complex cases or to provide lesser-experienced surgeons a way to practice (1.c.).

9.4.2 Healthcare Systems Engineering

As previously mentioned, healthcare is an extremely expensive sector. One contributing factor is the inefficiencies of healthcare processes and resources used. Digital twins have the power to reduce inefficiencies, for example by letting hospital staff simulate how to best organize their resources for their emergency department (2.a.). A hospital cannot simply adjust factors like staffing, room planning, and patient admissions without disrupting real-world operations. However, a digital twin of a hospital's emergency department allows the administrators to simulate scenarios, test solutions and optimize resource allocations. Once an ideal solution has been found using the digital twin, the changes (for example, the ideal timing to re-stock supplies, during which hours additional staff should be on call, or which medical devices are underused and may be moved to another department) can be implemented in the real-life emergency department [18]. These changes can help reduce hospital costs without compromising patient care. Furthermore, digital twins can optimize processes: [9], have successfully developed and deployed digital twins for trauma management (2.b.).

9.4.3 Health Monitoring and Remote Care

As mentioned multiple times in previous (sub)sections, with digital twins, patient data is continuously collected, analyzed, and used for predictions. This monitoring of the patient can be done in the hospital during an acute injury or illness, where the focus might lie in predicting the treatment and healing journey. Or it can be done remotely [2]. This remote monitoring can be ideal for the following scenarios: the patient is currently healthy, but faces an increased risk of developing a certain conidition or illness (for example due to genetic factors). The patient is living with a chronic condition, but is currently stable. The continuous remote monitoring in these cases ensures convenience and high quality of life for the patient, because he can carry on his life normally without missing work or other responsibilities to attend frequent health check-ups. It also reduces healthcare costs. At the same time, any concerning signs are detected early, which makes a successful intervention and outcome more likely (3.b.). Telemedicine is also a part of this type of care and has gained more popularity in recent years (3.a.).

9.4.4 Drug and Medical Devices Development

The process of drug discovery and development is a lengthy and costly one. It often takes years of trial and error, and still, only a small percentage of drugs make it to market. This is where digital twins come in, offering a way to revolutionize this process. Digital twins in drug discovery create virtual representations of biological systems, from individual cells to entire organs. These twins can simulate how a drug interacts with a system, predict side effects, and even test combinations of drugs - all virtually (4.a., 4.b.). This means fewer resources spent on physical experiments and faster identification of promising treatments [8], [3]. Digital twins can also help with life-saving medical devices, such as a pacemaker or surgical implant. Using digital twin technology, the effectiveness and safety of such medical devices are thoroughly tested before they ever touch a patient (4.c.). This reduces risks for patients and can shorten the time to market for new and innovative medical devices [26].

9.5 Digital Twin Regulatory and Ethical Considerations

There are many different regulatory and ethical considerations that must be taken into account if digital twins are to transform healthcare in the coming years and have a positive

impact on many, not just the institutions creating them or a few privileged individuals. The following subsections highlight some of the key challenges. Figures 9.1, 9.2, 9.3 provides an overview of all the papers analyzed, and the columns "key regulatory / ethical aspects" and "challenges of digital twins in healthcare" list the corresponding findings.

9.5.1 Data Privacy and Security

Privacy and security is a prominent challenge for many new technologies. Digital twins increase this difficulty due to their additional layers of data integration. Furthermore, they possess sensitive health data that must be protected from breaches, unauthorized access, and misuse. This requires adequate cybersecurity measures, which are often very costly [2].

9.5.2 Patient Autonomy and Data Ownership

Another challenge is the question of ownership and data rights. In the United States for example, current legal frameworks give the researchers or institutions that collect and work with the data ownership of it. This implies that the patients might not have ownership of their own digital twin, raising concerns that institutions instead of individuals would be in control. The potential for institutions to use digital twins in ways that individuals may not approve of adds another layer of complexity to this ethical dilemma [4].

9.5.3 Algorithmic Transparency

Algorithmic transparency poses a great challenge for the successful integration and widespread implementation of digital twins in healthcare. Algorithmic transparency is defined as the ability of stakeholders to understand and audit the decision-making processes of the underlying algorithms. In healthcare, this is especially important, because patients and doctors need to be able to trust digital twin technologies to make the right and fair decisions. But in the worst case, algorithms could actually exacerbate existing biases or errors in healthcare, and have a negative effect on the outcomes of patient and healthcare equity and accessibility. Because of the "black box" nature of advanced AI techniques, such as deep learning, [19] advocate for the integration of explainable AI techniques to make the decision-making processes of digital twins more interpretable. They also suggest enhancing transparency by including clinicians in the decision-making process and establishing industry-wide standards for algorithmic transparency.

9.5.4 Healthcare Inequality

Healthcare inequality is currently a pressing issue, and if digital twin technologies remain exclusive to privileged groups, it could further increase this inequality and create "digital divides". Therefore, policymakers must prioritize inclusivity and equitable access [10]. But this is challenging to do, because the technological infrastructure needed is not evenly distributed across the world or even across regions in a single country. Furthermore, digital twins are currently extremely expensive and further research funding is difficult to get, even in wealthy nations.

9.5.5 Integration, Scalability and Standardization

Digital twin deployment is costly and very complex. Digital twins require extensive integration of diverse data sources, such as personal health tracking devices, electronic health records, and environmental sensors. If data integration is insufficient, the result is data

silos that provide incomplete or fragmented datasets [10]. Furthermore, interoperability issues with existing healthcare systems must be considered, and even if these issues are solved, desired scalability of digital twins in healthcare is not given at the moment [5]. To facilitate scalability and widespread adoption of digital twins in healthcare, frameworks for standardization must be created. Standards such as ISO/IEEE 11073 already exist for personal health devices, but they are insufficient for digital twins [3].

9.5.6 Other Concerns

Other concerns include the issue of liability and accountability. If an incorrect prediction or recommendation is made by a digital twin, and the resulting treatment (or lack thereof) harms the patient, who should be held accountable? The institution that created the digital twin? Or the patient, who might have signed a waiver before the digital twin was created? These are all questions that are not answered yet, and hinder the adoption of digital twins in healthcare. Furthermore, new forms of discrimination or stigmatization, (particularly for individuals flagged as high-risk for certain conditions) might arise due to the powerful predictive capabilities of digital twins. Tying back to patient autonomy, patients should have the right to live without being constantly monitored, assessed and labeled, if they wish to do so. Some patients might not want to know if they have an incurable disease, but rather live day by day. But if their digital twin has this information about the patients' predicted health status, health insurance or other stakeholders might try and gain access to avoid paying for high costs in the future. Even if individuals remain healthy, the psychological and social impact of constantly being monitored might negatively impact their mental health and result in constant anxiety or unease.

The environmental impact of digital twins, especially when applied at scale, is also being criticized and should be kept in mind when creating new frameworks and regulations.

9.6 Future Research and Limitations

Despite the promising potential of digital twins in healthcare, several gaps and challenges remain and should be the focus of further research efforts. Digital twins rely on collaboration across many fields - engineering, healthcare, data science, economics, and more [14]. This makes the process complicated and costly. Future research should focus on fostering and supporting interdisciplinary work. This could include frameworks, specific recommendations or guidelines on how the collaboration can be executed, and relevant governmental entities included.

As discussed in 9.5, challenges can be technical (algorithmic transparency, data privacy and security, integration, scalability, standardization), ethical (patient autonomy, healthcare inequality, liability) or regulatory in nature. But arguably the most deciding factor of all is the economic factor. Although digital twins offer significant potential for improving healthcare outcomes, their high cost poses a barrier to widespread adoption. Future studies should conduct detailed cost-benefit analyses to quantify the economic feasibility of digital twins in various healthcare scenarios and explore strategies for reducing costs.

Long-term studies evaluating the impact of digital twins on healthcare outcomes, patient satisfaction, and economic efficiency are needed to ensure their safety and allow patients, doctors and other stakeholders to gain trust.

This seminar paper has several limitations that should be acknowledged. The literature review focuses on studies published between 2021 and 2024. While this ensures that the paper addresses recent advancements, it may overlook foundational research or earlier works that remain relevant. Due to the breadth of the topic, this paper emphasizes enabling technologies and economic implications. It does not provide an exhaustive

analysis of technical implementations or detailed case studies, which could offer additional insights. Furthermore, the paper lacks quantitative analysis, such as specific cost savings or return-on-investment (ROI) metrics, which would strengthen its economic arguments. This was out of scope for this seminar, but should be researched further if possible.

9.7 Conclusion

The healthcare sector is an extremely complex and expensive sector worldwide, facing challenges such as aging populations, increasing rates of chronic diseases, and limited resources. Despite these rising costs, the patient care and resulting outcome is not always satisfactory. Health issues are often diagnosed too late and treatment is usually done in a "one-size-fits-all" approach, despite the uniqueness of each patient. Digital twins represent a transformative opportunity to address these challenges by leveraging technologies like the internet of things (IoT), artificial intelligence (AI), and advanced analytics. By creating real-time, high-fidelity virtual replicas of physical systems, (ranging from individual organs to entire healthcare facilities), digital twins enable proactive patient monitoring, personalized treatment plans, and more efficient resource allocation. This shift from reactive to proactive healthcare has the potential to significantly improve patient outcomes while optimizing costs.

This paper conducted a systematic literature analysis, comparing 25 studies across multiple dimensions to provide a comprehensive overview of the current state of digital twins in healthcare. Key findings include the enabling technologies such as IoT for data acquisition, AI for predictive analytics, and blockchain for secure data handling, which were discussed in detail in section 9.3. Application areas like personalized medicine, surgical simulation, and drug development, highlighted in section 9.4, underscore the versatility and broad impact of digital twins. Additionally, the regulatory and ethical challenges, including data privacy, algorithmic transparency, and healthcare inequality, were explored in Section 9.5, emphasizing the need for responsible development and implementation.

While the potential of digital twins is vast, their adoption faces significant barriers. High implementation costs, integration complexities, and the lack of standardized frameworks are ongoing challenges. Moreover, ethical concerns around patient data ownership, accessibility, and equitable deployment must be addressed to ensure these technologies benefit all stakeholders, not just a privileged few. Overcoming these obstacles will require interdisciplinary collaboration, robust policy frameworks, and continued research to refine the economic and technical feasibility of digital twins.

Looking ahead, digital twins have the potential to revolutionize healthcare by bridging technological innovation with economic and societal needs. By enabling personalized, efficient, and equitable care, digital twins can transform the way healthcare is delivered, making it more sustainable and patient-centered.

Appendix A

List A (relevant)

1. Digital twins for healthcare 4.0-Recent advances, architecture, and open challenges
2. Digital twins in healthcare: is it the beginning of a new era of evidence-based medicine? A critical review
3. Leveraging digital twins for healthcare systems engineering
4. The use of digital twins in healthcare: socio-ethical benefits and socio-ethical risks
5. A comprehensive survey of digital twins in healthcare in the era of metaverse
6. Generative artificial intelligence empowers digital twins in drug discovery and clinical trials
7. Digital twins in healthcare: Methodological challenges and opportunities
8. Digital twins, internet of things and mobile medicine: a review of current platforms to support smart healthcare
9. Digital twins in healthcare: Applications, technologies, simulations, and future trends
10. Enhancing Healthcare through Sensor-Enabled Digital Twins in Smart Environments: A Comprehensive Analysis
11. Future outlook on the materialisation, expectations and implementation of Digital Twins in healthcare
12. Digital twins in healthcare and biomedicine
13. Concepts and applications of digital twins in healthcare and medicine
14. Digital Twins for Healthcare Using Wearables
15. The Rise of Digital Twins in Healthcare: A Mapping of the Research Landscape
16. Revolutionizing Healthcare: A Review Unveiling the Transformative Power of Digital Twins
17. The Economics of Digital Twins.
18. Increasing acceptance of AI-generated digital twins through clinical trial applications
19. AI and Digital Twins Transforming Healthcare IoT

20. Representation of Machine Learning Models to Enhance Simulation Capabilities Within Digital Twins in Personalized Healthcare

21. In Their Own Image: Ethical Implications of the Rise of Digital Twins/Clones/Simulacra in Healthcare

Appendix B

List B (specific examples/case studies)

1. The role of AI for developing digital twins in healthcare: The case of cancer care
2. Metaverse and healthcare: Machine learning-enabled digital twins of cancer
3. Digital twins in healthcare: an architectural proposal and its application in a social distancing case study
4. Integrating mechanism-based modeling with biomedical imaging to build practical digital twins for clinical oncology
5. Applications of digital twins in the healthcare industry: case review of an IoT-enabled remote technology in dentistry
6. TWIN-GPT: Digital Twins for Clinical Trials via Large Language Model
7. A framework for the generation of digital twins of cardiac electrophysiology from clinical 12-leads ECGs
8. Systems-based digital twins to help characterize clinical dose-response and propose predictive biomarkers in a phase I study of bispecific antibody, mosunetuzumab...
9. Cloud-based digital twins' storage in emergency healthcare
10. The emerging role of artificial intelligence and digital twins in pre-clinical molecular imaging
11. Towards a novel framework for reinforcing cybersecurity using digital twins in IoT-based healthcare applications
12. Digital twins elucidate critical role of Tscm in clinical persistence of TCR-engineered cell therapy
13. Building Digital Twins for Cardiovascular Health: From Principles to Clinical Impact
14. Healthcare for the Elderly With Digital Twins
15. SynTwin: A graph-based approach for predicting clinical outcomes using digital twins derived from synthetic patients
16. FMCPNN in digital twins smart healthcare
17. IDRes: Identity-Based Respiration Monitoring System for Digital Twins Enabled Healthcare

18. Unlocking the potentials of digital twins for optimal healthcare delivery in Africa
19. Cardiac Healthcare Digital Twins Supported by Artificial Intelligence-Based Algorithms and Extended Reality-A Systematic Review
20. A Self-Powered Dual Ratchet Angle Sensing System for Digital Twins and Smart Healthcare
21. Digital Twins in Healthcare and Their Applicability in Rhinology: A Narrative Review
22. Model-based digital twins of medicine dispensers for healthcare IoT applications
23. Advancing Healthcare Diagnostics: Machine Learning-Driven Digital Twins for Precise Brain Tumor and Breast Cancer Assessment
24. The Potential of Digital Twins in Healthcare: Evaluation of a Clinical Decision Support System for Chronic Inflammatory Bowel Disease
25. A virtual community healthcare framework in metaverse enabled by digital twins
26. A Framework of Digital Twins for Improving Respiratory Health and Healthcare Measures
27. Increasing the power in randomised clinical trials using digital twins
28. Digital twins for personalized healthcare: application to radiopharmaceutical therapies
29. Health Digital Twins with Clinical Decision Support and Medical Imaging
30. Toward Digital Twins for Human-in-the-loop Systems: A Framework for Workload Management and Burnout Prevention in Healthcare Systems
31. Optimize Healthcare Workflows: Sleeping Disorders Diagnosis and Challenges Using Digital Twins
32. Geospatial Information Based Digital Twins for Healthcare
33. Virtual Clinical Trials of BMP4 Differentiation Therapy: Digital Twins to Aid Glioblastoma Trial Design
34. Multisensor data fusion in Digital Twins for smart healthcare
35. A New Regulatory Road in Clinical Trials: Digital Twins
36. Digital Twins in Healthcare for Citizens MOSTLY ETHICAL ASPECTS, BACKUP LIST
37. Enhancing Longitudinal Clinical Trial Efficiency with Digital Twins and Prognostic Covariate-Adjusted Mixed Models for Repeated Measures (PROCOVA-MMRM)
38. Digital Twins in Human Activity Prediction on Gait Using Extreme Gradient Boosting Local Binary Pattern: Healthcare 6.0
39. Clinical application of virtual antiarrhythmic drug test using digital twins in patients who recurred atrial fibrillation after catheter ablation

40. Towards a Generation of Digital Twins in Healthcare of Ischaemic and Haemorrhagic Stroke

41. Evaluating digital twins for alzheimer's disease using data from a completed Phase 2 clinical trial

42. Pre-clinical imaging with artificial intelligence and digital twins

43. Abstract B023: Modeling of new drugs clinical trials outcome with patients' digital twins cohorts

44. The Use of Digital Healthcare Twins in Early-Phase Clinical Trials: Opportunities, Challenges, and Applications

45. The use of digital twins in healthcare: socio-ethical benefits and socio-ethical risks BACKUP LIST

Appendix C

List C (Discarded Papers)

1. Digital Twins and Healthcare: Trends, Techniques, and Challenges (NOT AVAILABLE)
2. Digital Twins in Healthcare: A Global Perspective on Adoption Trends, Challenges, and Impacts across Public and Private Hospital (NOT YET REVIEWED)
3. Digital twins and healthcare: Quick overview and human-centric perspectives
4. Virtual patients, digital twins and causal disease models: paving the ground for in silico clinical trials
5. Human Digital Twins for Pervasive Healthcare: A Scoping Review
6. Hierarchical federated learning based anomaly detection using digital twins for smart healthcare
7. Position paper From the digital twins in healthcare to the Virtual Human Twin: a moon-shot project for digital health research
8. Detecting latent topics and trends of digital twins in healthcare: A structural topic model-based systematic review (CONSIDERS MANY OLDER PAPERS)
9. How healthcare systems engineering can benefit from digital twins?
10. Deep learning-empowered clinical big data analytics in healthcare digital twins
11. Featuring Digital Twins in Healthcare Information Systems (NOT AVAILABLE)
12. AI-Enabled Healthcare and Enhanced Computational Resource Management With Digital Twins Into Task Offloading Strategies
13. Digital twins in healthcare: an assessment of technological and practical prospects (UNAVAILABLE)
14. Digital twins environment simulation for testing healthcare IoT applications
15. Digital Twins in Healthcare: A Survey of Current Methods (LOW IMPACT FACTOR SCORE)
16. Healthcare in the Era of Digital Twins: Towards a Domain-Specific Taxonomy
17. Imagining digital twins in healthcare
18. Digital twins-based data fabric architecture to enhance data management in intelligent healthcare ecosystems

19. ClinicalGAN: powering patient monitoring in clinical trials with patient digital twins
20. Framing blockchain-integrated digital twins for emergent healthcare: a proof of concept
21. Digital Twins in Healthcare: A forefront for knowledge representation techniques
22. Hyperreal Patients. Digital Twins as Simulacra and their impact on clinical heuristics
23. CONNECTED: leveraging digital twins and personal knowledge graphs in healthcare digitalization
24. Experts' View on the Future Outlook on the Materialization, Expectations and Implementation of Digital Twins in Healthcare (DUPLICATE)
25. Digital Twins-Empowered Secure Network Slice Access and Isolation for Consumer Healthcare Applications
26. Enhancing the Efficiency of Healthcare Facilities Management with Digital Twins
27. Digital Twins in Healthcare: Security, Privacy, Trust and Safety Challenges
28. Exploring Power Dynamics of Healthcare Digital Twins at Home through the Matrix of Domination
29. Challenges and Innovations in the Creation of Digital Twins in Healthcare
30. The Role of Digital Twins for improving Sustainability in Healthcare: The IRST Case
31. Real-time digital twins end-to-end multi-branch object detection with feature level selection for healthcare
32. Unveiling the Future: Blockchain-Powered Digital Twins for Personalized Privacy Preservation in Metaverse Healthcare Data
33. Digital Twins in Healthcare System: Communication between Society and Law
34. Cognitive Digital Twins for Improving Security in IT-OT Enabled Healthcare Applications
35. Digital twins and cybersecurity in healthcare systems
36. Explainability and the Role of Digital Twins in Personalized Medicine and Healthcare Optimization
37. RWD146 Multidimensional Analysis of the Implementation and Impact of Digital Twins in Healthcare
38. Towards Digital Twins in Healthcare: How would a meaningful Digital Twin for the user look like?
39. Digital Twins for Healthcare and Telecommunications Applications: A Survey
40. Elevating Precision Medicine: Uniting Digital Twins and AI in Healthcare Web-Service Platform Design

41. Resource Optimization with Digital Twins Using Intelligent Techniques for Smart Healthcare Management
42. Machine learning based models for implementing digital twins in healthcare industry
43. A Security Assurance Profile for IoT Digital Twins in Healthcare
44. Digital Twins for Proactive and Personalized Healthcare-Challenges and Opportunities (BOOK)
45. From the digital twins in healthcare to the Virtual Human Twin: a moon-shot project for digital health research
46. ...It Will Take to Cross the Valley of Death: Translational Systems Biology, "True" Precision Medicine, Medical Digital Twins, Artificial Intelligence and In Silico Clinical...
47. Digital twins will revolutionise healthcare (COULD NOT ACCESS)
48. The state of the art of digital twins in healthcare
49. Digital Twins in Healthcare: Proactive Regulation to Prevent a "Runaway Train" (MOSTLY LEGAL ASPECTS)
50. Digital twins and their appliance in transport economics
51. Digital twins in e-health: adoption of technology and challenges in the management of clinical systems
52. Advanced Technologies in Healthcare: AI, Signal Processing, Digital Twins and 5G (BOOK)
53. The case for digital twins in healthcare (BOOK, COULD NOT ACCESS)
54. Unlocking Potential: Proving the Value of Digital Twins to Healthcare Executives
55. Digital Twins in Healthcare: Addressing Concerns and Meeting Professional Needs
56. Digital twins in healthcare: State of the art and potential use cases in a hospital setting (FULL TEXT ONLY FRENCH)
57. Leveraging Data Physicalization in Healthcare Digital Twins: Enhancing Understanding and Interaction
58. Digital twins in healthcare (JUST A WORKSHOP OUTPUT)
59. A New Regulatory Road in Clinical Trials: Digital Twins: The promise and acceptance of this tool in study conduct is growing.
60. Empowering Intelligent Environments: Integrating Wearable Technologies and Digital Twins for Enhanced Healthcare and Well-Being
61. Appositeness of Digital Twins in Healthcare
62. Health Digital Twins with Clinical Decision Support
63. Corrigendum: CONNECTED: leveraging digital twins and personal knowledge graphs in healthcare digitalization
64. Digital twins and their appliance in transport economics

Table C.1: Abbreviations used throughout this paper and the figures.

Abbreviation	**Explanation**
5G, 6G	Fifth/Sixth -Generation Technology: Provides high-speed, low-latency connectivity for digital twins.
AI	Artificial Intelligence: Core to digital twin insights and advanced analytics.
AR	Augmented Reality: Overlays digital information on real-world environments for enhanced visualization.
BIM	Building Information Models: Models physical infrastructures to optimize workflows.
CNN	Convolutional Neural Networks: Analyzes image data, such as scans and diagnostic imagery.
CPS	Cyber-Physical Systems: Integrates computational systems with physical processes for real-time monitoring.
FL	Federated Learning: Enables model training on distributed data without centralizing sensitive patient information.
GANs	Generative Adversarial Networks: Creates synthetic data for training models in rare conditions.
GIS	Geographic Information Systems: Analyzes and visualizes geographic and spatial data.
IoT	Internet of Things: Enables real-time data collection through wearables, sensors, and medical devices.
KG	Knowledge Graphs: Organizes relationships between data points for easier querying and analysis.
LBS	Location-Based Services: Tracks spatial data for patient monitoring and navigation in healthcare facilities.
LSTM	Long Short-Term Memory: Processes sequential data like patient vitals over time.
MDI	Multimodal Data Integration: Combines diverse datasets (e.g., clinical, genomic, environmental) for holistic modeling.
ML	Machine Learning: Enhances analysis and predictive capabilities in digital twins.
MM	Mobile Medicine: Integrates mobile devices into digital twin workflows for remote monitoring and patient engagement.
Multi-Omics	Incorporates genomic, proteomic, and metabolomic data for comprehensive patient modeling.
PROCOVA	Prognostic Covariate Adjustment: Enhances predictions in clinical trials.
Robotics	Assists in surgical procedures and rehabilitation using digital twin models.
TL	Transfer Learning: Adapts pre-trained models for specific healthcare applications.
VAE	Variational Autoencoders: Compress and reconstruct data for accurate and efficient model creation.
VR	Virtual Reality: Provides immersive environments for training and surgery rehearsals.

Bibliography

[1] M. Abd Elaziz et al.: *Digital twins in healthcare: Applications, technologies, simulations, and future trends*; *WIREs Data Mining and Knowledge Discovery*, e1559, 2024, https://doi.org/10.1002/widm.1559.

[2] S. Adibi, A. Rajabifard, D. Shojaei, & N. Wickramasinghe: *Enhancing Healthcare through Sensor-Enabled Digital Twins in Smart Environments: A Comprehensive Analysis*; *Sensors*, 24(9), 2793, 2024, https://doi.org/10.3390/s24092793.

[3] M. Alazab et al.: *Digital Twins for Healthcare 4.0 - Recent Advances, Architecture, and Open Challenges*; *IEEE Consumer Electronics Magazine*, 12(6), 29-37, 2023, https://doi.org/10.1109/MCE.2022.3208986.

[4] B. Amram, U. Klempner, Y. Leibler, & D. Greenbaum: *In Their Own Image: Ethical Implications of the Rise of Digital Twins/Clones/Simulacra in Healthcare*; *The American Journal of Bioethics*, 23(9), 79-81, 2023, https://doi.org/10.1080/15265161.2023.2237456.

[5] P. Armeni et al.: *Digital Twins in Healthcare: Is It the Beginning of a New Era of Evidence-Based Medicine? A Critical Review*; *Journal of Personalized Medicine*, 12(8), 1255, 2022, https://doi.org/10.3390/jpm12081255.

[6] A. Balasubramanyam et al.: *Revolutionizing Healthcare: A Review Unveiling the Transformative Power of Digital Twins*; *IEEE Access*, 12, 69652-69676, 2024, https://doi.org/10.1109/ACCESS.2024.3399744.

[7] M. Batty: *Digital twins*; *Environment and Planning B: Urban Analytics and City Science*, 45(5), 817-820, 2018, https://doi.org/10.1177/2399808318796416.

[8] M. Bordukova et al.: *Generative artificial intelligence empowers digital twins in drug discovery and clinical trials*; *Expert Opinion on Drug Discovery*, 19(1), 33-42, 2024, https://doi.org/10.1080/17460441.2023.2273839.

[9] A. Croatti, M. Gabellini, S. Montagna, & A. Ricci: *On the Integration of Agents and Digital Twins in Healthcare*; *Journal of Medical Systems*, 44(9), 161, 2020, https://doi.org/10.1007/s10916-020-01623-5.

[10] C. De Maeyer & P. Markopoulos: *Future outlook on the materialisation, expectations and implementation of Digital Twins in healthcare*; *34th British HCI Conference*, July 2021, https://doi.org/10.14236/ewic/HCI2021.18.

[11] Z. Johnson & M. J. Saikia: *Digital Twins for Healthcare Using Wearables*; *Bioengineering*, 11(6), 606, 2024, https://doi.org/10.3390/bioengineering11060606.

[12] G. Korovin: *Digital Twins in the Industry: Maturity, Functions, Effects*; In V. Kumar, J. Leng, V. Akberdina, & E. Kuzmin (Eds.), *Digital Transformation in Industry*, Vol. 54, Springer International Publishing, 2022, https://doi.org/10.1007/978-3-030-94617-3_1.

[13] N. Kshetri: *The Economics of Digital Twins*; *Computer*, 54(4), 86-90, 2021, https://doi.org/10.1109/MC.2021.3055683.

[14] S. M. Kuriakose et al.: *The Rise of Digital Twins in Healthcare: A Mapping of the Research Landscape*; *Cureus*, 2024, https://doi.org/10.7759/cureus.65358.

[15] A. Marfoglia et al.: *Representation of Machine Learning Models to Enhance Simulation Capabilities Within Digital Twins in Personalized Healthcare*; *2024 IEEE International Conference on Pervasive Computing and Communications Workshops*, 100-105, https://doi.org/10.1109/PerComWorkshops59983.2024.10502444.

[16] C. Meijer, H.-W. Uh, & S. El Bouhaddani: *Digital Twins in Healthcare: Methodological Challenges and Opportunities*; *Journal of Personalized Medicine*, 13(10), 1522, 2023, https://doi.org/10.3390/jpm13101522.

[17] S. Mihai et al.: *Digital Twins: A Survey on Enabling Technologies, Challenges, Trends and Future Prospects*; *IEEE Communications Surveys & Tutorials*, 24(4), 2255-2291, 2022, https://doi.org/10.1109/COMST.2022.3208773.

[18] N. Mohamed et al.: *Leveraging Digital Twins for Healthcare Systems Engineering*; *IEEE Access*, 11, 69841-69853, 2023, https://doi.org/10.1109/ACCESS.2023.3292119.

[19] E. O. Popa et al.: *The use of digital twins in healthcare: Socio-ethical benefits and socio-ethical risks*; *Life Sciences, Society and Policy*, 17(1), 6, 2021, https://doi.org/10.1186/s40504-021-00113-x.

[20] K. Sel, D. Osman, F. Zare, S. Masoumi Shahrbabak, L. Brattain, J. O. Hahn, & R. Jafari: *Building digital twins for cardiovascular health: From principles to clinical impact*; *Journal of the American Heart Association*, 13(19), e031981, 2024, https://doi.org/10.1161/JAHA.123.031981.

[21] V. Sharma, K. Sharma, & A. Kumar: *AI and Digital Twins Transforming Healthcare IoT*; *2024 14th International Conference on Cloud Computing, Data Science & Engineering*, 6-11, https://doi.org/10.1109/Confluence60223.2024.10463366.

[22] A. Subasi & M. E. Subasi: *Digital twins in healthcare and biomedicine*; In *Artificial Intelligence, Big Data, Blockchain and 5G for the Digital Transformation of the Healthcare Industry*, Elsevier, 2024, https://doi.org/10.1016/B978-0-443-21598-8.00011-7.

[23] M. Turab & S. Jamil: *A Comprehensive Survey of Digital Twins in Healthcare in the Era of Metaverse*; *BioMed Informatics*, 3(3), 563-584, 2023, https://doi.org/10.3390/biomedinformatics3030039.

[24] A. A. Vidovszky et al.: *Increasing acceptance of AI-generated digital twins through clinical trial applications*; *Clinical and Translational Science*, 17(7), e13897, 2024, https://doi.org/10.1111/cts.13897.

[25] I. Volkov et al.: *Digital Twins, Internet of Things and Mobile Medicine: A Review of Current Platforms to Support Smart Healthcare*; *Programming and Computer Software*, 47(8), 578-590, 2021, https://doi.org/10.1134/S0361768821080284.

[26] K. Zhang et al.: *Concepts and applications of digital twins in healthcare and medicine*; *Patterns*, 5(8), 101028, 2024, https://doi.org/10.1016/j.patter.2024.101028.