

Identifying Anomalies in MFA with AI

Dissertation Report



April 5, 2022

Student Name

Student ID

Table of Contents

[Abstract 2](#_Toc100164500)

[Introduction 2](#_Toc100164501)

[Literature Review 2](#_Toc100164502)

[Risk-based Authentication 3](#_Toc100164503)

[Biometric Approach to MFA 6](#_Toc100164504)

[Implementation 6](#_Toc100164505)

[The 2FA Playground 6](#_Toc100164506)

[Backend Server 7](#_Toc100164507)

[Machine Learning Model 9](#_Toc100164508)

[Results, Analysis and Evaluation 9](#_Toc100164509)

[Conclusion 9](#_Toc100164510)

[Recommendations 9](#_Toc100164511)

[References 9](#_Toc100164512)

# Abstract

# Introduction

With the advent of technology, the security and safety of our electronic lives are not far-fetched. The only problem is that the development in security technology is a double-edged sword due to the freedom of access people have to technology. Thus, hackers also get their hands on this technology easily and find ways to get past it. Two-factor authentication (2FA) is one of the most innovative security measures to prevent hackers and malice from approaching our social accounts. 2FA requires a third-party app or device to verify the account details during login.

Cell phones assume significant parts in numerous individuals' day-by-day life. Individuals usually use cell phone applications to take photographs, send messages, book rides, or shop on the internet. It is not strange for those applications to ask for private data (like names, sex, or Visa data) from their clients to improve the nature of their administration. The delicate idea of those private data requires application engineers appropriately tie-down admittance to their administration. A mainstream approach to getting such access is by asking for passwords from clients during login measures (Irvan et al., 2021).

In any case, passwords and other information-based validation strategies like PIN (individual ID number) codes convey extraordinary danger as clients will in general utilize similar passwords across different administrations. Accordingly, numerous administrations right now require extra belonging based authentication techniques before allowing access. An ordinary method of this execution is by sending an interesting code through SMS (short message administration) to clients' telephone numbers. This additional progression is 2-factor validation (2FA) or multifaceted confirmation (MFA). Tragically, ownership-based validation techniques carry possible bothers to clients since e they may need to convey extra gadgets, which can be effectively lost. Numerous clients additionally utilize the same cell phone to enter passwords and get 2FA codes. In this manner, if their cell phone goes missing, assailants can sidestep 2FA checks (Irvan et al., 2021).

To prevent this, a novel solution of creating a third layer in the 2FA that will detect anomalous attempts using machine learning has taken its bearings for implementation. This also claims the title 3FA or the 3-Factor Authentication for ease of discussion in this paper.

In April 2019, Kaspersky researchers revealed an enormous scope of SIM trade misrepresentation activities focusing on clients in both the Portuguese-talking countries of Brazil and Mozambique had the option to utilize social designing, pay off, and straightforward phishing assaults eventually taking cash from casualties. Danger entertainers did these assaults by assuming responsibility for a casualty's telephone number by capturing accounts and catching two figure confirmation strategies in which the subsequent authentication factor is an SMS message or a call put to the portable number.

Two-factor validation, the additional security step that requires individuals to enter a code shipped off their telephone or email, has generally attempted to protect usernames and passwords from phishing assaults. In any case, security specialists have shown a mechanized phishing assault that can slice through that additional layer of safety—likewise called 2FA—conceivably fooling clueless clients into sharing their private qualifications.

Cases like these are popping up all over the country. Various hackers have compromised the integrity of the 2FA security. Researchers are trying to keep up with their security measures but the requisite of something intelligent and robust still exists.

# Literature Review

Authentication holds a mandatory place in the Cybersecurity Domain (Townsend, 2021). With the advent of MFA (Multi-factor Authentication), security has improved rapidly. However, access to new technology is not limited to the good people only. Cybercriminals are constantly evolving their methods and strategies including adding Artificial Intelligence to their list of tools (Townsend, 2021). Authentication is an early and basic line of safeguard for business information. Yet, traditional authentication because of passwords stays a flimsy part. That is because clients are famous for awful password practices (Townsend, 2021):

* Reusing passwords
* Utilizing unsurprising ones
* Putting away secret key data on tacky notes or in decoded bookkeeping sheets

Subsequently, an ever-increasing number of organizations are adding MFA (Multi-factor Authentication), in any event, for their client encounters, requiring clients to make an item buy or go through with a bank exchange from their telephone for additional authentication. By adding extra factors past the secret word or, far and away superior, instead of the secret key organizations can assist with defeating secret word shower and social designing assaults and stop hackers utilizing taken certifications from truly entering the record. Extra factors could incorporate addressing a security question, utilizing a one-time secret word, or answering a pop-up message on the telephone (Townsend, 2021).

Hackers are shrewd, however, and even with MFA, individuals' records might break and leak (Townsend, 2021). Gadgets, similar to telephones or USBs, are prone to stealing. OTPs (One-time Passwords) communicated through SMS are easily captured. What is more, biometrics, like fingerprints and, surprisingly, facial acknowledgement, can be hacked or faked. As Artificial, intelligence gets forward movement; it turns out to be considerably more straightforward to counterfeit even biometrics, making counterfeit fingerprints and facial pictures with an adequate number of matching focuses to pass an output. However, MFA improvement is possible by adding a basic snippet of data: context. Context is the data about the client's login, similar to where the client is while endeavouring to sign in or the gadget used. Such context can give basic insights that an assault is occurring (Townsend, 2021).

## Risk-based Authentication

To add context, the Identity and Access Management (IAM) industry has answered with risk-based authentication. Standard MFA catches data about what the client knows, similar to a password, what the client has, similar to their telephone, and even who the client is utilizing biometrics like fingerprints. Risk-based validation takes into account extra factors that help decide whether the client truly is who they say they are (Townsend, 2021). This is finished by contrasting their past login conduct with the present authentication endeavour, giving setting data that is absent in standard MFA. For instance, assuming a client ordinarily signs in on a specific PC from the fundamental office area during the week but abruptly attempts to sign in from a telephone at Starbucks, it could be an indication of a taken PC or compromised account.

On the other hand, assuming the client commonly signs in from home through one IP address and out of nowhere is signing in from another IP address (Townsend, 2021). Maybe one on a rundown of dubious IPs - you would need to challenge the login endeavour and request an extra validating component like a one-time secret word or face examination from a symbolic gadget. Evaluating authentication information like this continuously requires escalated and complex handling. That is where Artificial Intelligence comes in that place.

To execute risk-based authentication, cybersecurity organizations use AI-upheld innovations. The AI surveys and weighs individual variables about the login endeavour to think of a gamble score for the situation. For instance, a client associating with specific IP addresses or endeavouring to sign in during the centre of the night could show a danger (Townsend, 2021).

Artificial intelligence can likewise involve brain networks as a component of AI frameworks. These brain networks emulate the human mind and "learn" by being taken care of by datasets that incorporate the right outcome (Townsend, 2021). For instance, information about signing in utilizing various IPs and the outcomes demonstrating which of those logins were cyberattacks. It resembles a child offered a variable based math issue and the response, who should sort out what the recipe is to take care of this kind of variable based math issue. The AI grows endlessly better calculations to figure out which variables demonstrate an assault by attempting various methods to take care of the issue and looking at its response against the response in the dataset (Townsend, 2021). In the end, it tracks down a bunch of calculations that help anticipate dangers more often than not.

Risk engines screen various elements in a client's logins over the long run and assemble a profile for every client to comprehend login designs. At the point when a client differs from that profile on a given validation endeavour, the AI framework surveys the variable factors and decides a gamble score for the current login endeavour. A portion of the variables normally represented includes (Townsend, 2021):

* Network reputation
* Client's geographic area
* The device fingerprint (like the producer, model, or program)
* Login Time

While the vital advantage of AI-fueled risk-based confirmation is security, it can likewise smooth out the validation cycle. In standard MFA, clients are provoked for extra factors at each login endeavour. Enter your username and secret key, and then, at that point, answer a security question. Then again, enter your username and secret key, and then answer a message pop-up on your telephone. With AI-fueled confirmation, clients at generally safe probably will not be requested any extra factors, making login quicker (Townsend, 2021).

Risk-based authentication will proceed to improve and get more brilliant. At last, risk-based authentication will probably move from directed realizing, where the dataset incorporates the results, to solo realizing where the AI observes new examples that people might not have found and makes forecasts of expected variables to survey (Townsend, 2021). Having the option to cross-reference different AI calculations and use design acknowledgement and time-series, based prescient calculations will work on the precision and extent of AI-based validation contributions going ahead, for internet application logins, yet in addition for different parts of online protection like organization interruption and botnet recognition.

Simultaneously, designers will be searching for ways of giving IT divisions more command over the AI framework, for example, the capacity to see precisely why the AI settled on a given choice, change the number of perceived variables, and tailor the framework to their association's special climate. Albeit not stringently AI, different cross-industry drives are in progress to empower better information sharing so the data one association has on a potential danger can be made accessible to different associations progressively, further developing MFA (Townsend, 2021).

You can likewise hope to see AI-fueled validation frameworks extend to envelop consistent confirmation. Rather than constant danger evaluation exactly at login, AI frameworks will recognize and answer dangers all through a client meeting (Townsend, 2021). Assuming the client unexpectedly moves to another area and gadget, or endeavours to get to monetary data that is not pertinent to their work, the prompt to confirm their personality comes up.

In addition, surprisingly, further, access to the board will probably move from the application level to the information level. Specialists are now looking at appending metadata to individual bits of information, to demonstrate who ought to have what sort of admittance to that discrete snippet of data. For instance, the field in a data set containing worker compensations would include metadata demonstrating that main clients inside the organization who hold specific jobs can see that data (Townsend, 2021). At the point when that compensation data is exposed, the limitations on access arise. Artificial intelligence controlled validation would then implement these information level access limitations in any place the information is utilized.

As the character and access to the executives’ necessities advance, so will AI as an instrument in IAM. Since, the truth of the matter is, AI is important to deal with the intricacy of investigation at the scale and speed that will be required in the changing danger scene and the developing character and access to the board climate (Townsend, 2021).

MFA utilizes any mix of at least two elements to confirm personality and keep crucial resources secure from fake access. At this point, we've all pre-owned two-factor validation (2FA) online to approve a login or exchange by joining a secret word with an SMS code shipped off our cell phone. In the event of the compromise of a component, the framework is yet secure (Kightlinger, 2019).

Three primary elements take the place in play to affirm character (Kightlinger, 2019):

* Something you have – a physical item, for example, an ATM card, key fob or USB stick.
* Something you know - "confidential" like a password or PIN.
* Something you are - a biometric element, for example; fingerprints or voice, iris scans and other physical attributes.

The guidelines for how to consolidate these elements and use them to verify character relies upon the element carrying out them. In specific businesses, Current expectations from MFA require it to meet consistency commands. For instance, the Payment Card Industry Data Security Standard (PCI DSS) requires MFA for character and access to the executives in unambiguous conditions - for example, remote admittance to a cardholder information climate that begins from outside the organization or administrator admittance to the information climate from inside the confided in the network (Kightlinger, 2019).

An ever-increasing number of associations are thinking about consistently on MFA for each application and IT framework, yet all the same, that is quite often excessively lumbering. Assuming representatives need to hang tight for an SMS code to arrive at their telephone each time they need to get to an application, client purchases decrease with time (Kightlinger, 2019). A more compelling way to deal with getting the venture includes strategies to lessen the weight of activity on the client however much as could be expected and focusing on the applications that require 2FA in light of awareness and chance of giving and take.

Indeed, even where MFA is justified, nonintrusive gamble based or context-oriented confirmation can make it less baffling for clients. Nonintrusive validation factors incorporate gadget fingerprinting, geolocation, IP, gadget reputation, and portable organization administrator information (Kightlinger, 2019). Some danger insight stages, for example, the IBM X-Force Exchange, as of now give this data to outsider applications and arrangements.

These components add the setting to the client and gadget for an exchange and assist with evaluating the gamble level of every activity. If the gamble is too extraordinary, extra confirmation is required. For example, assuming a client in New York signs in to the corporate organization utilizing her work area, you may not need MFA; yet if a client in Hong Kong attempts to get to an application through an obscure organization utilizing an unnoticed gadget, you certainly need to add validation measures (Kightlinger, 2019).

Stages incorporating fraud detection technologies and unified endpoint management (UEM) tools assist with diminishing the requirement for client-driven MFA and give accommodating settings about the client's gamble level to decide the requirement for extra layers of verification (Kightlinger, 2019). Such stages engage associations to oversee and get every one of the numerous ways representatives to an interface when they are portable, for example, cell phones, PCs, wearables and even web of things (IoT) gadgets. An open stage likewise makes coordination with existing applications and foundation direct.

MFA might be fine for workers, who can be expected to utilize anything confirmation instrument their association picks. Organizations have customarily weighed security versus accommodation, continuously underscoring the last option out of dread that clients would dismiss additional means to safeguard individual information (Kightlinger, 2019).

In any case, this customary way of thinking may presently not be valid as we are seeing an expanded degree of acknowledgement and experience with multifaceted confirmation from everyone. Truth be told, as the "IBM Future of Identity Study 2018" showed, shoppers have become more natural and tolerant of MFA, particularly concerning cash related applications and virtual entertainment. Contingent on the age bunch, the sort of MFA favoured differs, with the more youthful age substantially more OK with cell phone innovation and biometric techniques or tokens as opposed to passwords (Kightlinger, 2019).

Organizations could observe that the ideal arrangement is to give clients a decision among different verification choices, whether that is one-time passwords or fingerprint readers (Kightlinger, 2019). Risk-based approaches similar to those for employees become usable in access scenarios for consumers. As the likely mischief from unusual movement rises, so can the number of confirmation factors required.

Techniques for MFA are constantly changing as weaknesses emerge; innovation advances and the predominant players progressively come from the millennial and Gen Z populaces (Kightlinger, 2019). New MFA approaches should supplant bulky logins with captivating, super-advanced conceivable outcomes. Brilliant organizations will remain adaptable and versatile by using a cloud stage that updates with the most recent techniques.

Moreover, dealing with picking an MFA strategy/technique is now like a piece of information-driven analysis - security pioneers ought to shift focus over to stages that permit them to screen the achievement paces of their verification techniques (Kightlinger, 2019). The strategies and arrangements you carry out today will not and ought not to be extremely durable. Gathering information persistently will assist you with formulating multifaceted validation techniques that give the ideal security, accommodation and complexity for your workers, clients and organization.

## Biometric Approach to MFA

To upgrade the security of corporate frameworks, for example, cell phones, security elements are conveying MFA to add an extra layer of insurance. MFA depends on three variables:

1. Something You Know (e.g., a secret key)
2. Something You Have (e.g., a symbolic gadget or Short Message Service (SMS))
3. Something You Are (e.g., fingerprints or voiceprints).

Indeed, even with the organization of MFA, it is yet conceivable to sidestep the MFA validation process. Consider the situation wherein there is a cell phone in possession (Something You Have). If security does not claim the protection of the cell phone and does not have a screen lock, the hoodlum can reset the secret phrase on financial applications, or even get client data from the gadget (Something You Know). A Phishing assault is one more method for taking a client's qualifications, where the casualty opens an email or message and clickbait fools the user into tapping on a noxious connection, which can prompt getting the Something You Have and Know. Indeed, even dependence on (Something You Are) is risky. For instance, one can utilize a general fingerprint (genuine or engineered fingerprints that can serendipitously coordinate with an enormous number of fingerprints [5]) which can break 65% of genuine fingerprints [6], or one can utilize the casualty's biometric examining, for example, a fingerprint, voice recording or photograph, which can be acquired from the gadget. To stay away from the recently referenced Phishing assaults, an individual's uniqueness has been utilized as verification apparatus. A client's mark is an instrument utilized for client verification and has been in need for a long time. Even more as of late, marks have arisen as a biometric acknowledgement apparatus. The client needs to give their unmistakable, which drives them to put away information in the framework. Nonetheless, if the put away biometric information were to be compromised, the information would be significant to the hacker's hands (e.g., fashioning marked documents can be utilized). This drove us to ponder a more secure method for joining the client's interesting credits to concede admittance to delicate information on their cell phones.

# Implementation

The project has two basic modules. The modules are as follows:

## The 2FA Playground

The 2FA playground will be, as the name suggests, a playground for testing and experimenting in this project. The playground will be enabled with two main features i.e., the 2FA auth. and request tracking.

The programming language used for this module will be JavaScript and the tools employed from JS will be the **Node.js** and **express** server libraries to create the backend server that the app will run from and the **speakeasy** library to implement the 2FA. The **speakeasy** library functions combined with the **Google Authenticator browser extension** to generate tokens that verify the user Login process. The main functions as mentioned previously for this module will be to provide an API that allows 2FA user registration, user authentication and user validation. In addition, this module will be responsible for collecting client request data from the requests on this module. User data like the **IPv4 address, IPv6 address, email and password** are targets for the collection so that they can train the Machine Learning Model to identify anomalies.

This module will use the free Heroku server for hosting. This will make the module online and accessible by different devices having different IP addresses. This will allow us to collect data from different devices under different scenarios also providing a much-needed variation in the data. The information of the user freshly registered will be stored as JSON in an onboard JavaScript object whose access will be to all the project files to avoid any access issues down the road.

### Backend Server

The backend server whose function will be to provide a roadmap for all the above processes to link together will be coded using the express library in JavaScript. Since this application will only be a proof of concept, there will be no special server hosted databases and proper user authentication. The application will automatically generate a random user ID and save it along with a secret that is specific and unique to each user. The following code is responsible for this:

app.post("/api/register", (*req*, *res*) *=>* {

*const* id = uuid.v4();

  try {

*const* path = `/user/${id}`;

*const* tempSecret = speakeasy.generateSecret();

    db.push(path, { id, tempSecret });

    res.json({ id, secret: tempSecret.base32 });

  } catch (error) {

    console.log(error);

    res.status(500).json({

      message: "Error Generating the Secret!",

    });

  }

}); // Register User in DB + Create Temporary Secret for each User

The express server in JavaScript provides access to different functionalities by practicing a technique called routing where each function has a different URL associated to it. The above code snippet shows that the user registration has its route ‘**/api/register’** reserved to it. Since this is a proof of concept, the server will be running locally on a local port specified by us, Port: 5000 in this case. Therefore, the URL for user registration is **‘localhost:5000/api/register’**.

There is a function that runs when the URL is hit. In the function there is a try-catch block for exception/error handling. The function uses the speakeasy library to generate a secure and unique secret for each user and save it along with a uniquely generated user ID using the UUID library in the database. The database is stored locally as a file in the local directory of the code itself. The database stores all its information as JSON. This following is a snapshot of the database:

{

*"user"*: {

*"44bf50a3-e23d-4b33-b960-2306f7909d66"*: {

*"id"*: "44bf50a3-e23d-4b33-b960-2306f7909d66",

*"secret"*: {

*"ascii"*: "8sBHDZ]sgbG(I8t)\*rTxUSztZw[?>2xT",

*"hex"*: "38734248445a5d7367624728493874292a72547855537a745a775b3f3e327854",

*"base32"*: "HBZUESCELJOXGZ3CI4UESODUFEVHEVDYKVJXU5C2O5NT6PRSPBKA",

*"otpauth\_url"*: "otpauth://totp/SecretKey?secret=HBZUESCELJOXGZ3CI4UESODUFEVHEVDYKVJXU5C2O5NT6PRSPBKA"

      }

    }

  }

}

As you can see, the user data is stored as a JSON object in which there is a secret and unique key generated for each user in 3 different formats namely, **ascii, hex** and **base32.** We will only be requiring the base32 format for this although generating all the three formats is crucial to the security aspect.

The next step will be to verify the user upon login using real 2FA. For this research, we will be employing 2FA using a 3rd party authenticator app, more specifically **Google Authenticator**. For this research, I use the Google Authenticator extension available for most popular browsers. The verification code is as follows:

app.post("/api/verify", (*req*, *res*) *=>* {

*const* { userId, token } = req.body;

  try {

*const* path = `/user/${userId}`;

*const* user = db.getData(path);

*const* { base32: secret } = user.tempSecret;

*const* verified = speakeasy.totp.verify({

      secret,

      encoding: "base32",

      token,

    });

    if (verified) {

      db.push(path, { id: userId, secret: user.tempSecret });

      res.json({ verified: true });

    } else {

      res.json({ verified: false });

    }

  } catch (error) {

    console.log(error);

    res.status(500).json({

      message: "Error Finding the User!",

    });

  }

}); // Verifying the Token from Authenticator

The above code verifies the token against the unique secret previously generated for the user. Normally, the process of automatically generating a verification token by the Authenticator is app done automatically, but since this is not a full-fledged app, but only a playground, we add the user token in the authenticator app manually as follows:

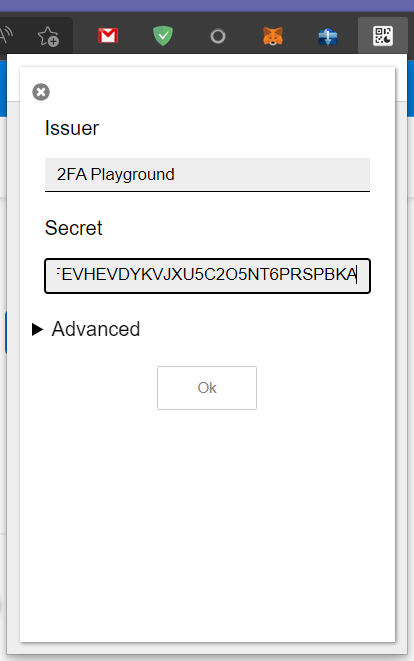


Figure 1 - Manually Adding User for 2FA Token Generation from 3rd Party Authenticator App.

As shown in *Figure 1,* You need to mention the issuer name and the secret for manual registration. The issuer name can be anything, preferably the username of the user. The secret will be the secret generated for each user uniquely but in base32 format. Once this is done, the authenticator will start generating tokens that expire after a few seconds. Copying one of these tokens and passing it onto the **“/api/verify”** route will return **verified = “true”** if the token has not expired and is correct and **verified = “false”** if the token is expired or wrong.

## Machine Learning Model

The Machine Learning Implementation requires further research to finalize it properly. The most crucial part of the implementation of Machine Learning part is the dataset. The dataset for training Algorithms on anomaly detection in 2FA is not available as of yet. Further research may uncover an existing dataset or the data collection will be the choice.

In case the dataset requires creating it from scratch, dummy data from the playground developed will provide the necessary input and anomalies doped and induced as a proof of concept for training the algorithm and providing results. Depending upon the features in the dataset (created or found online), the Machine Learning approach can vary from simple binary classification to Deep Learning or even unsupervised learning using clustering. Further research will help finalize these facts.

# Results, Analysis and Evaluation

# Conclusion

# Recommendations

# References

* Irvan, M., Thao, T., Kobayashi, R., Nakata, T. and Yamaguchi, R., 2021. Learning from Smartphone Location Data as Anomaly Detection for Behavioral Authentication through Deep Neuroevolution. Seventh International Conference on Information Systems Security and Privacy.
* Ali, G., Ally Dida, M. and Elikana Sam, A., 2020. Two-Factor Authentication Scheme for Mobile Money: A Review of Threat Models and Countermeasures. Future Internet, 12(10), p.160.
* Anderson, W. and Simske, S., 2020. At-Home Healthcare Through Smart-Environmental Sensing, Including Biometrics for Multi-Factor Authentication. 2020 IEEE International Conference on Healthcare Informatics (ICHI).
* Asani, E., Longe, O., Balla, A., Ogundokun, R. and Adeniyi, E., 2020. Secure Human-Computer Interaction. Handbook of Research on the Role of Human Factors in IT Project Management, pp.149-163.
* Iyer, A., Karthikeyan, D., Hasan Khan, M. and Binu, D., 2020. AN ANALYSIS OF ARTIFICIAL INTELLIGENCE IN BIOMETRICS-THE NEXT LEVEL OF SECURITY. Journal of critical reviews, 7(1).
* Karim Abdul-Hassan, A. and Hasson Hadi, I., 2019. INTELLIGENT AUTHENTICATION FOR IDENTITY AND ACCESS MANAGEMENT: A REVIEW PAPER. Iraqi Journal for Computers and informatics, 45(1), pp.6-10.
* Kesem, L., 2021. AI Hand Pattern Authentication Method. Boise State University.
* Kightlinger, D., 2019. Beyond 2FA: Secure Your Critical Assets with Risk-Based Multifactor Authentication. [Online] Security Intelligence. Available at: <https://securityintelligence.com/posts/beyond-2fa-secure-your-critical-assets-with-risk-based-multifactor-authentication/> [Accessed 5 April 2022].
* Kwon, B., Sharma, P. and Park, J., 2019. CCTV-Based Multi-Factor Authentication System. Journal of Information Processing Systems, 15(4), pp.904-919.
* Townsend, A., 2021. AI's Role in Authentication. [Online] OneLogin. Available at: <https://www.onelogin.com/blog/ai-authentication> [Accessed 5 April 2022].