Identifying the Invisible:
A Comprehensive Approach to Distinguishing Software Bots

Zhixiong Chen
Math & Computer Science Dept.
Mercy University
Dobbs Ferry, New York, USA
email: zxchen@mercy.edu

Deming Chen
Electrical & Computer Engineering Dept.
University of Illinois Urbana Champaign
Champaign, Illinois, USA
email: dchen@uiuc.edu

***Abstract*— In the evolving digital landscape, software bots or bots have emerged as autonomous agents capable of engaging in complex interactions within computer-mediated environments. This research paper delves into the unique features and characteristics of bots, proposing a sophisticated framework for their identification and registration. These captured features are certainly different from those for human users. Central to our approach is the development of a holistic identification model that treats bots as integral components of social-technological ecosystems. By adopting a comprehensive methodology that includes the construction of portfolio artifacts, we aim to encapsulate both invariant identification characteristics and dynamic, verifiable credentials of bots. These artifacts serve not only as a means of distinction but also as a basis for ensuring security and authenticity in interactions involving them. Our work underscores the importance of a nuanced understanding of bots, advocating for a system that recognizes their potential while safeguarding against misuse. Through a meticulous analysis of bot behavior and interaction patterns, we contribute to the establishment of a more secure, transparent, and efficient digital environment where bots and human users coexist harmoniously.**

Keywords - Software Bot, Identification, Invariant, Registrar, Verifiable Credentials, Portfolio Artifacts.

I. INTRODUCTION

Bots, mingling with human users, have become an established practice in many social media applications [1]. Human users are gradually accepting bots not only as an inevitable technological development but also because they are increasingly humanized and intelligent [2][3]. Recently, Large Language Model (LLM) powered Artificial Intelligence (AI) assistants have demonstrated their capability of understanding human questions in depth and can provide comprehensible answers [4]-[6]. Furthermore, with techniques, such as downstream model fine-tuning, prompt tuning, Retrieval-Augmented Generation (RAG), and prompt engineering, these bots will become more understandable, intelligent, specialized, and eventually integrated into digital workforce. To prepare for this new reality, we need to have understanding to identify bots and to aggregate their timed verifiable credentials.

This position paper seeks to identify invariant characteristics of bots so we can use them to differentiate bots among themselves and also from human actors across a wide range of contexts. The goal is to treat a bot just like any other actor in social-technological applications, or akin to a worker in various workplaces. To achieve this accurately, we compiled a comprehensive glossary of vocabulary and special terminology specific to bot identification and provided clear definitions for each of them. This list will continue to expand as we gain a deeper understanding of the intrinsic nature of bots, their enabling technologies, and associated threat models.

The primary contribution of this paper is the definition of bot identity through various attributes. While no single attribute may uniquely identify a bot, a combination of them could do the work. We utilize similarity distances between bots for identification purposes. Additionally, we link bot invariant identity with verifiable credentials that evolve gradually, all recorded as Portfolio Artifacts (PAs). These PAs provide a holistic view on bot and would have an impact on bot development and registration.

Furthermore, we introduce 'Vital Plugins'—essential to bot functionality. These plugins function as standard APIs, enabling bots to perform a range of operations and services, including secure communications, to update Pas, and to response to requests for identification, authentication, and verification.

We employ ontology technology to ensure that PAs are well-defined within their schema, making them portable across different applications and machines. We design the format to be extendable, accommodating future developments and advancements, and scalable to meet the demands of an expanding digital workforce.

Our study emphasizes the crucial role of bot identification across a diverse range of applications, from education, social media to healthcare. This research highlights how effectively identifying bots can address challenges inherent in various sectors.

We organize our paper as follows. In Section II, we present literature review and related work. We brainstorm five mind maps in Section III. They include robots, botulation, botfession, botvaluaton, enabling technology, bot registrar, services, threat models, invariants, and applications,

providing a comprehensive classification and framework. In Section IV, we detail the representation of bot identity in a standardized format using JavaScript Object Notation for Linked Data (JSON-LD). This ensures we define all key attributes clearly within its schema. In Section V, we discuss the bot registration service. It employs blockchain technology to ensure data integrity and traceability. We conclude the paper with insightful discussions on both technological solutions and legal recommendations aimed at incentivizing the registration of bots. This includes a critical analysis of potential impacts and the benefits of formalizing bot identity in digital ecosystems.

II. Related Work

Recent research has significantly advanced the development of chatbots, particularly through the utilization of LLMs [6]-[11]. It is evident that LLM powered bots are becoming increasingly proficient at mimicking human conversation. However, literature searches focusing on invariant identity characteristics and dynamic, verifiable unique abilities of bots yield few results. Similarly, studies on bot threat models, including ethical considerations, are limited, with only a handful of papers mostly focused on bot detection [20]. One intriguing experiment noted that participants could only correctly identify the nature of other users—bot or human—42% of the time, even though they were aware of both bots and humans in the experiment. This indicates that there is still much to learn about identifying intrinsic bot features. Further, [13] observed that AI powered bot detectors could sanitize social media applications and enhance bot detection capabilities. IDPro's blog offered a brief list of suggestions for developing basic bot identity capabilities [7]. Additionally, a report at [22] suggests that to achieve more intuitive and adaptive learning capabilities, AI systems might benefit from a set of foundational behaviors, akin to biological CliffsNotes, which could serve as a form of inherited digital DNA.

Completely Automated Public Turing test to tell Computers and Humans Apart (CATCHA), is a type of challenge-response system used in computing to determine whether the user is human. It is a common security tool used on websites to prevent bot-based spam and automated data extraction. As AI technology evolves, CAPTCHAs have become obsolete to sophisticated bots capable of image recognition and text analysis. So, we believe such features, such as text, image, audio, math based CAPCHAs, are no longer distinguish software bots from human users.

III. Indentifications

We employ mind maps to brainstorm the terminology associated with bots and their interrelationships.

A. Bot

Figures 1-5 depict a bot ecosystem. In the right side of Figure 1, we list tangible side of bots, robots. We physically touch and feel them. The most advanced are 'synthetic robots' capable of reasoning, feeling, and consciousness, although they are shown only in fiction now. It is inspired by the synthetic beings featured in the TV series Humans [21]. The left side highlights distinct types of bots, collectively referred to as 'botulation'—a term we use to describe the bot population.

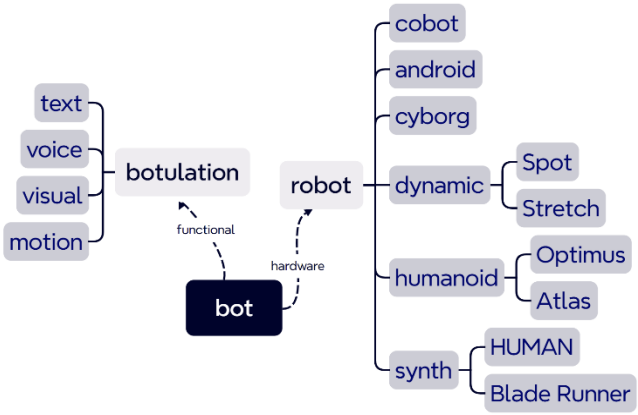

Figure 1. Bot Mind Map on Bot Population and Robot.

The right side of Figure 2 highlights the capabilities of bots within specific professional domains or sectors, which we use 'botfession,' anticipating their integration into the digital workforce. The left focuses on the metrics, tools, and techniques used to evaluate and assess bot's abilities, akin to methods used in grade school evaluations.

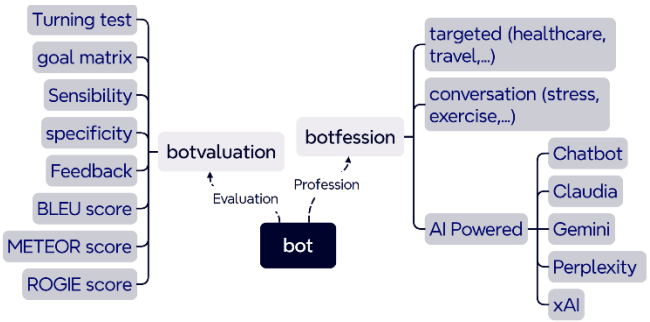

Figure 2. Bot Mind Map on Bot Profession and Evaluation.

Figure 3 delves into the enabling technologies for bots and registration authority. The left side explores the infrastructure that supports bot registration. This can be either centralized, similar to a certificate authority or a government agency, or decentralized platform, such as blockchain technology or any append-only storage. Given that our implementation utilizes blockchain, detailed insights into this technology are provided in the following sections.

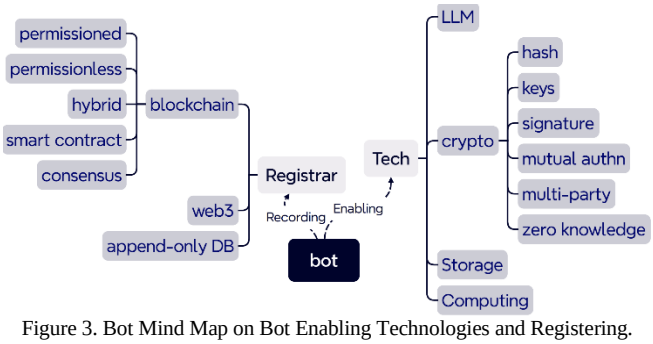

Figure 3. Bot Mind Map on Bot Enabling Technologies and Registering.

Figure 4 lists bot services and potential threat models that typically target human users. Bots are no longer considered categorically as bad actors. The right side presents a collection of operations and services essential for maintaining a bot's health. This includes bot registration, which is crucial for officially declaring a bot's existence. Updating is vital for fixing code bugs and continuous upgrades. Identification operations are crucial for verifying identity and assessing credentials, while secure communication ensures the use of secure protocols. Finally, interoperability is the key for automation of machine to machine. These operations are implemented as software plugins, which can be developed by third parties and invoked as needed. Collectively, these essential functions are referred to as 'vital plugins'.

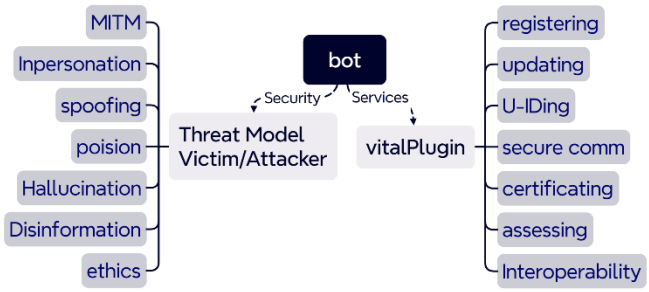

Figure 4. Bot Mind Map on Bot Services and Threat Model.

The right side of Figure 5 highlights invariants that can uniquely identify a bot from others. It is holistic that includes a variety of identifications, such as LLM models, codebase, training data (in other words, knowledge domain), configuration and portfolio. The left side lists examples bot applications.

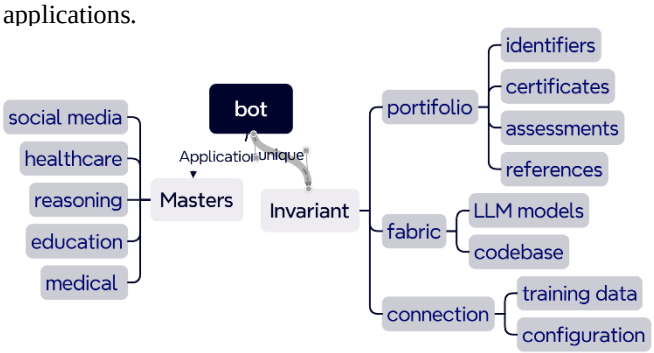

Figure 5. Bot Mind Map on Bot Invariants and Applications.

B. *Identification characteristics*

We can categorize bot identification into two types: assigned and inherited, analogous to Social Security Numbers (SSN) versus DNA. We begin by exploring the assigned features.

- UID Group

This group encompasses identities, such as given names, universal IDs, and domain professions. The Universal ID is structured as a sequence of bits separated by dots, like an IP address. Each segment represents a level of sub-grouping, allowing for specific sub-group information to be retrieved through masks.

- Crypto Group

This group includes a unique pair of public and private keys for each bot, which are essential for unique identification, secure communication, authentication, and integration. The key pair is generated using a randomized algorithm. To prevent man-in-the-middle attacks, the public key should be registered in a publicly verifiable domain, which could be managed by a trusted third party like a Certificate Authority (CA) or stored on a decentralized, immutable blockchain. An Initialization Vector (IV) may be introduced to protect against replay attacks. The private key remains secured within the bot and shall never be transmitted externally.

- Master

This group records information about the bot's owner, revealing the intended use of the bot, its service duration, and other pertinent details. It includes built-in control mechanisms to ensure that the bot adheres to the expectations set by its master. Mathematically, it reflects applications.

We now turn to the extraction of inherited features of a bot, which are indicative of its development process.

- Code Base and Configuration

This group identifies the bot's code base, incorporating elements such as a code hash or checksum, referred to as a 'tag,' to detect any unauthorized alterations. This tag is digitally stamped, ensuring code assurance and security—vital for maintaining a healthy bot ecosystem in the digital workforce. The code version maintains a log of changes and updates, reflecting the bot's evolution over time. Additionally, more granular aspects of the code base, such as the code structure, token sequences, and distributions, provide deeper insights into the bot's internal structure beyond what a tag can offer. We propose using LangChain to achieve this detailed analysis [23].

Configuration holds equal importance as it details the training processes of the bot and the operational parameters under which it was utilized. Together, the codebase and its configuration form a comprehensive bot signature.

- Credentials

Credentials consist of one or more claims made by an issuer. A verifiable credential is a tamper-evident credential whose authorship can be cryptographically verified [16][17]. This category is unique in that it not only reflects a bot's capabilities but also its achievements over time. It includes specialized skills, underlying AI architecture (e.g., rule-based, retrieval-based, deep neural network, or hybrid), performance metrics to date, and areas of continuous learning.

Bot metrics also serve as unique identifiers, derived from consistent measures, such as the bot's latency, vocabulary size, and conversational patterns, some of which are detailed in the 'Botvaluation' in Figure 2.

Additional credentials, such as licenses issued to the bot by various platforms, can further enrich this profile.

C. *Identification and Security*

- Bot Identification

Bot identification is a critical process designed to verify a bot's identity, ideally without human intervention. This process is facilitated by a software plugin, referred to as

'VitalPlugin,' embedded into the bot's code. It includes a standard API that processes identification requests among other functions. Using cryptographic tools, the plugin enables secure communication—a common practice in social media applications. We propose designing specific protocols for handshakes and identity verification, adhering to the principle of 'always ask, never assume.' This approach ensures that the bot can autonomously confirm its identity and operate securely.

- Mitigating Security Threats

To counteract potential attacks identified in the threat models of Figure 1, we employ advanced cryptographic techniques, including digital signatures, multi-layered protection, and unique challenge-response sets that only the bot can generate. Protecting the private key is paramount to prevent unauthorized access and impersonation.

- Enhancing Verification Measures

While basic identifying information can be spoofed, true verification requires additional authenticity signals.

a) Registry Systems: A centralized registry could officially list verified bot identities and credentials, making it difficult for impersonators to falsify these details.
b) Multi-factor Authentication: Bots could be required to periodically re-verify their identities through multiple authentication factors, such as keys, signatures, and one-time codes.
c) Platform Verification: Platforms on which bots operate could provide verification indicators, such as a verified checkmark, to confirm the authenticity of the bots.
d) Behavioral Analysis: Analyzing and comparing the conversational patterns and tendencies of an original bot against those of a potential impersonator can help detect deviations in interaction styles.
e) Code Analysis: Ensuring the integrity of a bot's codebase involves verifying that the actual code of an impersonator matches the identified hashes and expected programming standards.
f) Social Graph Analysis: Examining the social connections and interaction patterns of an original bot versus an impersonator can reveal significant differences, aiding in the detection of fraudulent entities.

IV. REPRESENTATION OF BOT IDENTITY

We utilize PAs to represent the identity attributes of bots [18][19]. PAs serve as fundamental components for identifying not only human beings but also bots and other applications. In this context, our focus is specifically on bots. We employ JSON-LD to outline the ontology design, which can be visualized using tools like the JSON-LD Playground. JSON-LD leverages URLs, @context, and IRIs to enhance semantics, extensibility, and interoperability. It allows for the linking of entities using IRIs instead of internal indices and supports integration with digital signatures and proofs. Proof mechanisms, such as LD Signatures, enable the signing of JSON-LD documents.

Figure 6 highlights a simple example of a translation bot, although it does not present many features defined in Section 2. Figure 7 illustrates a segment specifically for identification purposes. They represent a single type of bot. The use of ontology and JSON-LD offers significant advantages in aggregating and reasoning about data.

```
{
 "@context": {
  "schema": "https://schema.org/",
  "Bot": "schema:SoftwareApplication",
  "feature": "schema:feature"
 },
 "@type": "Chatbot",
 "name": "ZC",
  "feature": [
  {
   "@type": "feature",
   "name": "Natural language processing",
   "description": "Processes and understands natural language input"
  },
  {
   "@type": "feature",
   "name": "Speech recognition",
   "description": "Converts spoken audio to text"
  },
  {
   "@type": "feature",
   "name": "Text-to-speech",
   "description": "Converts text to human-like speech"
  },
  {
   "@type": "feature",
   "name": "Conversational intelligence",
   "description": "Dialog management, context tracking, personalization"
  }
 ]
}
```

Figure 6. Bot Representation.

```
"issuer": "https://example.edu/issuers/14",
"issuanceDate": "2023-10-15T10:20:24Z",
"proof": {
 "type": "RsaSignature2018",
 "created": "2023-10-15T10:20:54Z",
 "proofPurpose": "assertionMethod",
 "verificationMethod": "https://example.edu/issuers/keys/1",
 "jws": "xxxxx"
}
```

Figure 7. Bot Identification.

A. PA Aggregation

PAs can be organized in a hierarchical format, analogous to branches in a body of knowledge. This structure allows for the aggregation of tags like <ComputerScience ∘ ProgrammingLanguage ∘ *> to demonstrate a bot's proficiency in various programming languages within computer science, such as Python, Rust, or Go. The aggregated claims are stored and communicated using descriptive rather than numerical identifiers from the PA tags, though actual execution relies on a numerical system for efficiency.

Aggregation is particularly useful when employing existing PAs to generate specialized credentials for tasks that require specific knowledge and skills, while deemphasizing other credentials.

B. PA Reasoner

We can also derive new claims using a PA Reasoner, which infers certain credentials from existing PAs. Our colleagues at Cyber Talent Bridge [24] developed a proof of concept for this approach. They utilized a rich set of linked data vocabularies published [e.g., 25] to enhance the Cyber Talent Bridge credentials. According to the open world principle, any vocabulary can be employed to extend the data of a Cyber Talent Bridge credential, which is also encoded in JSON-LD. JSON-LD facilitates the transformation of JSON documents to and from RDF serializations, such as N-Quads based on the specified @context.

V. IMPLEMENTATIONS

Introduced and refined since 2017, the Personal Archive Service System (PASS) and its enhanced iteration, PASS+, are built around blockchain technology [18]. The core of PASS uses PAs as foundational elements to establish unique digital identifiers for subjects, focusing particularly on non-human entities. It is used as a register platform for bot identification storage.

The PASS framework consists of an integrated system of modules, applications, and libraries. These components work collaboratively to facilitate the creation, storage, retrieval, and presentation of PAs. The structure of PASS is designed to leverage the decentralized and secure nature of blockchain, ensuring that each PA remains tamper-proof and verifiable. This architecture not only enhances the reliability of digital identities but also supports a scalable system for managing digital archives across various applications.

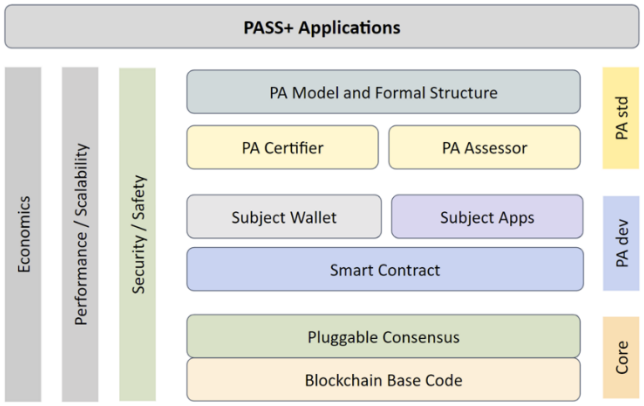

Figure 8. PASS Framework Layered Overview.

The architecture of the Personal Archive Service System (PASS) is depicted in Figure 5, showcasing a multi-layered framework. At its core, the system builds on blockchain technology, utilizing the Ethereum open-source code base along with pluggable consensus protocols. We have also implemented a permissioned blockchain using Hyperledger Fabric to meet specific security and governance requirements.

Above the core blockchain layer is the PA Standard (PA std) Layer, dedicated to the modeling and formal structure analysis of PAs. This layer ensures the systematic definition, classification, and organization of PAs. It features key modules, such as the PA Assessor and PA Certifier, which are crucial for evaluating and certifying the integrity and accuracy of PAs.

The PA Development (PA dev) Layer follows, focusing on the creation of decentralized applications. This layer includes modules for smart contracts, subject wallets, and web communication APIs that interface with other modules. The smart contract module is specifically designed for creating, storing, and retrieving PAs, ensuring that interactions with the blockchain are secure and efficient.

At the top of the architecture are the user applications that leverage PASS for various specific use scenarios, demonstrating the practical application of the system in real-world contexts.

Running parallel on the left side of the architecture are cross-layer considerations: security and safety, performance and scalability, and economic factors. These aspects are integral to the architecture, influencing every layer to ensure that the system remains robust, efficient, and economically viable.

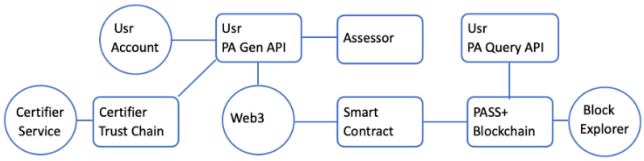

Figure 9. Service Ecosystem and Its Workflow.

Figure 9 presents an alternative view of the PASS system, specifically illustrating the workflow diagram. In this visualization, internal modules within the system are depicted as squares, while external applications or libraries are represented by circles, clearly distinguishing between PASS components and external integrations.

PASS interacts with bot applications primarily through the 'VitalPlugin', which utilizes a standard API to facilitate communication. This setup positions PASS as a decentralized registrar, a critical function in managing digital identities. We advocate that, particularly within social media platforms, this decentralized approach is preferable to traditional centralized authorities. This is due to its ability to enhance security, ensure greater privacy, and provide resilience against single points of failure.

This workflow diagram in Figure 9 not only highlights the structural and functional relationships within the PASS system but also underscores the system's adaptability and efficiency in handling real-world application demands in social media contexts.

VI. CONCLUSION AND FUTURE WORK

Registering bots and making their presence explicit in social media or any applications is challenging, particularly in environments like bot farms, which are often employed to influence public opinion. While technology has improved our ability to detect bots in social media applications—for instance, using conversational data to train bot detectors—these advancements also aid in the development of bots that can evade detection. This creates a continuous cycle of improvement in both AI-powered bots and AI-powered bot

detectors, complicating efforts to fully expose all bots. A viable approach to encouraging transparency is through legislation. With advancements in technology, such as the development of sophisticated LLMs, public acceptance of bots in social interactions is increasing, potentially reducing resistance to registration. However, the challenge remains significant. Legislation could be a powerful tool to mandate the registration of all bots.

Although threat modeling is a well-established research area when the focus is on human users, it is less frequently discussed in the context of bots. We believe that results from traditional threat modeling can be adapted to address threats posed by bots, enhancing security measures and protocols.

Several critical questions remain unanswered, including who controls the Bot Registrar, who develops and standardizes the VitalPlugin, and who is responsible for auditing and approving these systems. Additionally, what types of evaluation and assessment are necessary? How do we monitor a bot that evolves over time or potentially goes rogue? What metrics should we use to assess bot performance?

This paper outlines our recent research on bot identification and security, highlighting the complexity of managing bot in digital environments. The issues discussed require further investigation and the development of robust, standardized solutions to ensure the safe integration of bots into social media and beyond.

ACKNOWLEDGMENT

We would like to extend our gratitude to the various online resources and support tools that have provided invaluable guidance and feedback throughout the drafting of this paper. Additionally, this research was conducted under an appointment to the Summer Research Team Program for the U.S. Department of Homeland Security Science & Technology Directorate, Office of University Programs. We are thankful for the opportunity and support that have significantly contributed to the development of this work.

REFERENCES

[1] Z. Chen, Z. Lu, A. Sane, and A. Bhimsain, “Trustworthy When Human and Bots Are Mingled”, in Proceeding of the 7th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud), pp. 76-81. 2020.

[2] S. B. Kondracki and N. J. Nikiforakis, “Uninvited Guests: Analyzing the Identity and Behavior of Certificate Transparency Bots”, in Proceedings of the 31st USENIX Security Symposium, Security 2022, pp. 53-70, 2022.

[3] “The ultimate guide to machine-learning chatbots and conversational AI”, https://www.ibm.com/watson-advertising/thought-leadership/machine-learning-chatbot, last accessed 5/15/2024.

[4] A. Vaswani et al., “Attention Is All You Need”, v7, 2023, https://arxiv.org/abs/1706.03762, last accessed 5/15/2024.

[5] H. Touvron et al., “LLaMA: Open and Efficient Foundation Language Models”, https://arxiv.org/pdf/2302.13971.pdf, 2023, last accessed 5/15/2024.

[6] W. X. Zhao et al., “A Survey of Large Language Models”, https://arxiv.org/pdf/2303.18223.pdf, 2023, last accessed 5/15/2024.

[7] https://idpro.org/bot-identity/, last accessed 5/15/2024

[8] D. Banerjee, P. Singh, A. Avadhanam, and S. Srivastava, “Benchmarking LLM powered Chatbots: Methods and Metrics”, https://arxiv.org/pdf/2308.04624.pdf, 2023, last accessed 5/15/2024.

[9] S. Roller et al., “Recipes for building an open-domain chatbot”, in Proceedings of the 16th Conf. of the European Compt. Ling., https://aclanthology.org/2021.eacl-main.24/, last accessed 5/15/2024.

[10] Y. Li, S. Qu, J. Shen, S. Min, and Z. Yu, “Curriculum-Driven Edubot: A Framework for Developing Language Learning Chatbots Through Synthesizing Conversational Data”, in CHI’24, 2024.

[11] M. A. Kuhail, N. Alturki, S. Alramlawi, and K. Alhejori, “Interacting with educational chatbots: A systematic review”, Education and Information Technologies, V. 28, No 1, pp. 973–1018, 2023.

[12] Y. Bai et al., “Constitutional AI: Harmlessness from AI Feedback”, https://arxiv.org/abs/2212.08073, 2022 last accessed 5/15/2024.

[13] Davis University of California, “Gunrock 2.0: A user adaptive social conversational system”, in Alexa Prize SocialBot Grand Challenge 3.

[14] Amazon Science, https://www.amazon.science/alexa-prize/proceedings/gunrock-2-0-a-user-adaptive-social-conversational-system, last accessed 5/15/2024.

[15] K. C. Yang and F. Menczer, “Anatomy of an AI-powered malicious social botnet”, arXiv:2307.16336, 2023, last accessed 5/15/2024.

[16] W3C, “Decentralized Identifiers (DIDs)”, DID-CORE, https://www.w3.org/TR/did-core/, accessed 5/15/2024.

[17] W3C, “Verifiable Credentials Data Model v1.1”, Publication VC Data Model, https://github.com/w3c/vc-data-model/, https://w3c.github.io/vc-imp-guide/, 2023, last accessed 5/15/2024.

[18] Z. Chen and Y. Zhu. “Personal Archive Service System using Blockchain technology: Case study, Promising and Challenging”, in Proceeding of the 2017 IEEE International Conference on AI & Mobile Services, pp. 93-99, 2017.

[19] Y. Zhu and Z. Chen, “RealID: Building A Secure Anonymous Yet Transparent Immutable ID Service”, in Proceeding of the IEEE International Conference on Intelligent Data and Security, pp. 26-28, 2017.

[20] K. Radivojevic, N. Clark and P. Brenner, “LLMs Among Us: Generative AI Participating in Digital Discourse”, https://arxiv.org/abs/2402.07940, 2024, last accessed 5/15/2024.

[21] https://en.wikipedia.org/wiki/Humans_(TV_series), last accessed 5/15/2024.

[22] L. Zeldovich, “Why AI needs a genome”, https://www.cshl.edu/why-ai-needs-a-genome/, 2021, accessed 5/15/2024.

[23] LangChain, https://www.langchain.com/, accessed 5/15/2024.

[24] https://cybertalentbridge.com/, last accessed 5/15/2024.

[25] https://csrc.nist.gov/pubs/sp/800/181/r1/final, last accessed 5/15/2024.