Unifying Privacy and AI: Why I am an Entropist

Harry Halpin
Nym Technologies
harry@nymtech.net

***Abstract*—With the ascendance of artificial intelligence (AI), one of the largest problems facing privacy-enhancing technologies (PETs) is how they can successfully counter-act the large-scale surveillance that is required for the collection of data–and metadata–necessary for the training of AI models. While there has been a flurry of research into the foundations of AI, the field of privacy-enhancing technologies still appears to be a grab-bag of techniques without an overarching theoretical foundation. However, we will point to the potential unification of AI and PETS via the concepts of signal and noise, as formalized by information-theoretic metrics like entropy. We overview the concept of entropy ("noise") and its applications in both AI and PETs. For example, mixnets can be thought of as noise-generating networks, and so the inverse of neural networks. Then we defend the use of entropy as a metric to compare both different PETs, as well as both PETs and AI systems.**

***Keywords*—artificial intelligence, privacy-enhancing technologies, entropy, information theory**

I. Introduction

Our fundamental thesis is that privacy-enhancing technologies and artificial intelligence can be considered two sides of the same coin: Artificial intelligence (AI) systems are based on minimizing entropy in noisy data to discover information, while on the other hand privacy-enhancing technologies (PETs) can be characterized by the addition of noise (entropy) to obfuscate information. However, so far this relationship between AI and PETs has not even been sketched.

Although there was considerable interest in privacy after Snowden's revelations of mass surveillance, at the present moment both academic and popular attention are focused on AI. Yet AI is neither magic nor much of a philosophical wonder if one grasps its technological essence: The incredible feats that Large Language Models (LLMs) like ChatGPT are capable of are the result of processing vast amounts of human data in order to build generative probabilistic models. These models in turn can generate the very patterns that they originally learned from the underlying human-generated data, but customized to respond to user input [1].

In fact, the current gold rush in building AI systems may in fact prefigure a revival of PETs, for behind the curtain of AI's capabilities is a social quandary: The data which fuels these AI systems is collected via vast surveillance systems controlled by a few large companies, and often includes not just public data such as Wikipedia but also sensitive personal data [2]. Beyond waxing poetic about the future of superintelligence or panicking over the possible loss of jobs due to automation, one unrecognized challenge that AI presents is that in its quest for evermore sources of data, AI itself poses a threat to the continued existence of privacy and thus human autonomy [3]. However, privacy-enhancing technologies (PETs), such as anonymous communication systems and differential privacy, both lack deployment in practice and are fragmented in terms of research, and so are unable to impede the spread of AI-based surveillance.

While one practical reaction to the rise of AI is to immediately push PETs into deployment, this requires building new real-world systems that may take years. Yet how do we know if these PETs even work against powerful AI adversaries? So another less intuitive reaction to the challenge posed by AI is to revisit the lack of solid theoretical foundations for PETs. Only with some unified paradigm and clear metrics can (1) PETs be compared to each other and (2) the performance of PETs be measured against AI-based surveillance. As a starting point, we suggest a research paradigm based on information theory and entropy, rather than considering PETs to be a subfield of information security in some broadly conceived manner or just another form of applied cryptography. In this position paper, we informally sketch how the well-known paradigm of the discovery–and hiding–of a signal in noise is the common thread that unites both PETs and AI, and allows them to be measured in terms of entropy. This revival of what has been negatively termed *entropism* in a positive valence opens the space for PETs to systematically challenge the growing hegemony of AI [4].

The overarching goal of this position paper is to revive the concept of entropy (also considered as "noise" or "uncertainty" in terms popular parlance) in both AI and PETs as a core guiding principle of both fields. First, in Section II, we define AI as the general-purpose detection of information ("signal") within noise. In contrast, in Section III PETs are defined as the general-purpose application of the obfuscation of information by noise, leading to the concept of a *noise-generating networks* (NGN) as a counterpart to neural networks. This is exemplified by noise-generating mixnets (NGM) in the present literature [5]. We then confront Syverson's objections to entropy [4] in Section IV. In Section V, we outline future research directions in fleshing out this theoretical framework that has only been sketched so far.

Workshop on Innovation in Metadata Privacy: Analysis and Construction Techniques (IMPACT) 2025
28 February 2025, San Diego, CA, USA
ISBN 979-8-9919276-2-8
https://dx.doi.org/10.14722/impact.2025.23076
www.ndss-symposium.org

II. Artificial Intelligence

With enough complexity in their design, AI systems–perhaps more accurately termed "machine-learning" algorithms–are able to find signals (which can also be thought of as "patterns" or "regularities") across almost any form of data. In other words, what AI systems do is reduce uncertainty and so discover information within data.

In terms of information theory, AI can be thought of as non-linear function-fitting for arbitrary functions, where the function determines the signal and the rest of the data is the noise [6]. For an intuitive example: Imagine you are having a phone conversation while walking down a busy street: the voice on the other end is the signal, while the noise is everything else in the background. We often fail to hear particular words and have to guess their meaning using clues on the context: A signal is discovered via fitting despite all the noise based on our predictions of likely words due to our "training" in language. The signal leads to our understanding of speech via the correct prediction of missing aspects. In supervised learning, the nonlinear function is learned from the "training" data, and then extrapolated to new data, i.e. the "test" data. Neural networks can detect signal in existing data and predict functions in future data.

In many cases, signals may be so difficult to discover that no human mind can detect them of its own accord, even when given the powerful predictive theoretical frameworks and tools of modern science. In this case, AI systems are now stepping in to discover these hidden signals. While humans can be easily overwhelmed by the amount of noise–whether it's the sound of a subway interrupting a phone call or complex flows of traffic through a network–an AI can easily sift through noise with the patience and speed of its seemingly infinite data processing capability. The world has many hidden structures, and AI promises to discover them. These signals may not map to analytically or scientifically discoverable laws, or even signals that can be easily formalized via mathematics, but nonetheless can be captured and predicted via AI [7]).

In information theory, Shannon entropy measures the amount of randomness or uncertainty in a system. So a binary variable that has the maximum amount of entropy (1.0) is completely random, and any decrease in entropy can be thought of as a gain in information. In the far more common multivariate case, the entropy is bounded by $log_2(N)$ where N is the number of values that can be taken by a variable. In classical approaches to AI, entropy is often used to decide what feature partitions a data-set, such as in decision-trees, where relative entropy (i.e. reducing the entropy by the maximum amount given all other possible features) is used to choose the feature that "splits" the decision-tree optimally [8]. Turning to more sophisticated machine-learning algorithms, one rule to choose a prior probability distribution in Bayesian inference is to choose the distribution with the maximum entropy, thus the name "maximum entropy classifier" [9] However, in general Bayesian machine-learning with a clear foundation in information theory have fallen out of favor in AI (as well as simple techniques such as decision-trees, although they are valued for their ease of explanation in creating transparent and trustworthy AI systems), and to a large extent previous AI systems have been replaced by neural networks [1].

Today, AI is dominated by neural networks. Although AI is quite a varied field with many different types of models, neural networks are one of the most flexible models. Neural networks are composed of layers of artificial neurons, which in turn can each be considered to take part of the task in a distributed way, such as a classification or regression task. Each neuron can be considered a mathematical function that maps from one set of vectors to another set of vectors [10]. Working in combination, the precise kinds of multivariate functions can be implemented by a neural network depend on the number of layers and other parts of the structure of the neural network. Although neural networks were long derided as unprincipled, they are remarkably adept at learning, and new techniques such as deep learning enable representation learning by neural networks, where the network learns features from unlabeled data itself.

Although the original neural networks of a single layer fell out of fashion due their inability to capture even simple XOR functions [11], surprisingly neural networks with additional layers were able to capture increasingly complex functions, although their inner works remained – and still are – a relative mystery. The *universal approximation theorem* shows that a sequence of neural networks are in theory capable of capturing arbitrary functions for anything from image recognition to natural language if it can be captured by a mathematical function: For each function f, there exists a sequence of neural networks $\phi_1, \phi_2, \ldots$ such that $\phi_n \to f$ [11], a finding that was recently generalized to multivariate functions [12]. More importantly, it appears that with enough data for training and enough layers of artificial neurons, a neural network can discover any arbitrary hidden signal by learning from past data. Once trained, a neural network can discover the signal in future data. Typically, as the amount of data on which the network is trained increases, the more accurate it becomes at discriminating signal from noise [13]. This training is a "black box" as the universal approximation theorem simply states that, for a given function, a neural network exists in theory that can approximate the given function, but how to train a neural network to approximate a given function in reality is another matter entirely.

Just because a neural network is theorized to simulate a function in general, there is no guarantee that such a neural network exists, and the universal approximation theorem gives no guidance on the number of networks, layers, and other structure. Thus, in order to create such a neural network, it must be trained via some form of error-correction where their predictions are tested against some "ground truth" training data or via a policy for reinforcement learning when there is no training data. Similar to logistic regression, neural networks end up being built upon entropy: It appears most of the myriad cases of training can be reduced to minimizing entropy [14].

Neural networks often use cross-entropy or relative entropy, which attempts to minimize the divergence between the ground truth and predicted distributions. Of course, simplistic predictions would fail to generalize, so entropy can also be used to regularize the network so that the training does not overfit the data.

AI can also be used to *generate* new information, not simply classify or predict information. Generative AI like ChatGPT simply uses the underlying probabilistic function learned from training to generate new information, but as "perturbed" by a user query. This use-case of predicting and generating English language texts via building a language model was foreseen by Shannon himself in his early work in the 1950s, but with the rise of large amounts of data on the Web [15]. What happened recently to cause ChatGPT to increase in efficacy is that a new kind of more effective neural network sequence, the generative pre-trained transformer, was effectively trained on the entire Web [16]. Since the Web is really an aggregate of human information and interactions, AI began to discover patterns in what appears to be our entire social form of life. ChatGPT can detect signal in natural language questions via one neural network (an encoder) and use these signals to generate answers in text via another neural network (the decoder) [1]. Related AI algorithms can do the same with images, videos, and anything else imaginable, and various techniques that can also be formalized from the standpoint of entropy for multi-modal learning and generation [17].

III. Privacy Enhancing Technologies as Noise-Generating Networks

Yet there is also a dark side to AI. In the end, the training data comes from somewhere: it may be your personal life or secret information [16]. If so, AI systems can be used to predict and control your behavior with everything from "deepfakes" to persuading you to endorse a particular political view [11]. It can even be used to exterminate you via a drone strike! For example, an AI algorithm that is trained on your location data can discover your daily routine, and so can predict your future location. These kinds of uses of AI is cause for serious privacy concerns.

So how can we stop AI? The answer is simple: add noise to your information, such as blurring existing locations and creating fake locations. In effect, this can be thought of as using a form of AI against AI. We will argue that this concept of the addition of noise is at the heart of privacy-enhancing technologies, from anonymous communication networks to differential privacy.

On an historical aside, the formalization of the intuitive concepts of signal and noise in terms of entropy was invented by Shannon in his foundational paper that defined information theory [18]. What is not quite as well-known is that the motivation for Shannon's invention of information theory was his earlier cryptographic research that was classified at the time [19]. Information theory is to a large extent based on cryptography with the classified cryptographic elements removed and the concepts generalized.

Noise can be added at different levels of a system. In particular, cryptography can be defined as the production of entropy at the level of the information-carrying messages themselves. This entropy is created in a message (encryption) by the application of the entropy given in the key required to encrypt and decrypt a message (and various equivalents in terms of other constructions such as digital signatures). As defined by Shannon, perfect information-theoretic secrecy holds if $H(P|C) = H(P)$, where P is the possible values of the plaintext and C the possible values of the encrypted ciphertext. In terms of information theory, the ciphertext and the key used to encrypt the plaintext have the same amount of entropy. For example, noise can used as a one-time pad key for a message. Of course, most cryptographic systems are defined using weaker notions of secrecy, such as semantic security (where the adversary's advantage in inferring information is negligible over a given bound, rather than non-existent) and the weaker (and lesser-known) entropic security, a cryptographic scheme that guarantees that the entropy of a message is above a certain bound for a given adversary [20]. This application of information theory to cryptography is perhaps most clear in coding-based cryptographic systems such as the McEliece cryptosystem [21].

A larger problem is metadata, the information that can be inferred from the sending of messages. This often includes aspects traditionally left out cryptography, such as the length of a message or the time it was sent. More importantly, it includes the probability distribution of messages. On a higher-level of abstraction, metadata includes the distribution of information in general, such as the distribution of characteristics like gender in a database. Unlike personal data and classified information, metadata has almost no legal protections in regulations like the General Data Protection Regulation (GDPR), making metadata ripe for abuse. Cryptography can defend the privacy of the information carried by messages, but PETs are needed to defend the privacy of metadata. Given the amount of actors that can observe metadata, PETs should be a thriving area of research.

Yet PETs is a relatively small field of research despite its importance. Unlike cryptography, which has been given a systematic formal information-theoretic treatment by Shannon that led to various notions of provable security [19], privacy-enhancing technologies lack a unified conceptual framework based in information theory and so have a more fragmented formal foundation than cryptography, although various subfields of PETs around particular technologies such as anonymous communication networks and differential privacy have rigorous formal foundations. In general, entropy seems to best fit the bill for creating both a unified foundation and a comparative metric, as entropy is used in privacy-enhancing technologies in use-cases ranging from anonymous communication [22] to location privacy [23].

Superficially, what we see in privacy-enhancing technologies is the rise of many ad-hoc measurements, with over 80 distinct measurements being recorded in a survey of privacy metrics [24]. Perhaps the original metric is the simple size is

the anonymity set, which measures how many other individuals could fulfill a particular function. Being uniquely identified usually means the reverse of privacy (or anonymity), and has been given new terms such as *unicity* in the literature [25].

Differential privacy seems quite similar to entropy-based approaches to privacy, as it uses noise to achieve privacy in databases. Indeed, differential privacy does explicitly refer to entropy in making its guarantees [26]. So differential privacy can naturally be interpreted in terms of entropy. Yet this unity may not be immediately apparent, as differential privacy comes with a bewildering zoo of measurements: k-anonymity that measures the number of other individuals that can be distinguished by a query [26], l-diversity that measures the variation in certain attributes and t-closeness measures whether or not a certain attribute's distribution in the results of a query matches that of the entire database [27] More confusingly, new metrics are constantly proposed such as k-map, δ-presence, β-likeness and δ-disclosure. Thus, there is a real risk that each new technology, or even each new experiment, produces its own metric for success. Yet the core metrics such as k-anonymity can each in their respective foundational paper be measured in terms of entropy [26]. What matters is often 'what' is being measured, and this may be easier to measure using a derivative metric, but the core metric of entropy remains valuable.

From an intuitive standpoint, what is done in privacy-enhancing technologies is typically to obscure information via noise (deleting attributes or individuals from a database) or add new noise to the dataset (adding fake individuals to a database). Although most research on anonymous communication systems has focused on mixing [28], most modern mixnet systems actually add "fake' or "dummy traffic," although seemingly as an afterthought [5]. However, traffic can be added in a non-uniform manner, as to match bursty or power-law distributions during internet usage. Regardless of the particulars, it is precisely this ability to add noise that is essential, and mixing and cryptography itself can be thought of as variants as noise. Therefore, we can generalize a sort of *anti-neural network* that adds noise (and thus entropy) to prevent information discovery by a neural network. One useful way to think about this is a *noise generating network* (NGN): To invert the universal approximation theorem [11] to a universal obfuscation theorem, for each function f, there exists a sequence of noise generating networks that can add noise $\delta_1, \delta_2, \dots$ to neural networks $\phi_1 + \delta_1, \phi_2 + \delta_2 \dots$ such that there does not exist a neural network $\phi \to f$. Noise-generating networks allow us to conceptualize differential privacy and mixnets as doing the same abstract function of adding entropy to a distribution.

A stratified mix network with fully connected layers of mix nodes appears–both graphically and formally in terms of an ordered sequence of fully connected layers of nodes–to be quite similar to a neural network, but with a different purpose. So mixnets can be thought of as a kind of noise-generating network, in particular a *noise-generating mixnet* (NGM). For example, when one sends traffic through the Nym NGM, cover traffic is also sent out with your "real" data packets [5]. From the outside, these packets look exactly like real ones, but they're empty. This not only obscures the type of content you're sending–for instance, by making a data transfer during some period look larger than it is–but also increases the overall anonymity of the network for everyone. Adding delays can also be thought of as adding noise to an underlying distribution by changing the frequency, rather than adding or subtracting packets. Thus, data mixing creates timing obfuscation so that the order and frequency of packets handled by a node are scrambled and cannot be analyzed to reveal the traffic patterns of users, for instance, based on when a packet arrives and leaves from a server. In the future, noise-generating mixnets should be built for scaling. Surveillance systems are now global, so privacy technology needs to be capable of scaling to meet the demand as internet traffic increases.

IV. Why I am an Entropist

The most famous objection to the use of entropy to study anonymous communications is Paul Syverson's polemic "Why I am not an Entropist" [4]. Syverson is particularly noteworthy as he is one of the original inventors of *onion routing* and Tor [29], who despite his background in philosophy ended up working at the Naval Research Laboratory and procuring the original funding for Tor. Thus, due to the widespread success of Tor, the use of entropy-based metrics has fallen out of favor as a metric for anonymous communication systems. We will argue against his point that entropy gives information to the adversary or is impossible to estimate. In general, it appears Syverson's argument is not against entropy per se, but against the all-knowing global passive adversary (GPA) that can track each and every packet's flow throughout the network, an adversary that just happens to be beyond the scope of Tor's threat model [29].

Syverson argues informally, much in the same style as this paper, that entropy should not be the guiding metric for research into anonymous communication system, much less privacy more widely as we have argued. Although Syverson does note that the problem "is not that entropy entirely fails to reflect uncertainty," Syverson argues against *entropism*, which he characterizes as "using entropy as the meaning, the criterion for anonymity or how anonymous something is as it "does not capture everything important" [4]. Of course, no metric captures everything, and even a metric such as entropy has numerous variants (Renyi and Hartley, as noted by Shannon) and so the choice of a particular measure of entropy depends both on the data and the problem at hand. It is self-evident to note that no quantitative metric is perfect or even necessarily sufficient in of itself for any particular problem. Yet lacking a unified metric that allows comparison between different techniques leads research to be incomparable, lost in a quagmire of ad-hoc measurements, where each particular technique can brandish its own unique measurement to claim success.

Syverson then claims that entropy "provides your adversary with an explicit target that he has available resources to over-

come" [4]. While the observation that a measurement gives an adversary some vital clue in deploying their finite resources may have been valuable in the 1990s, today computational resources are increasingly cheap, and so it should not be assumed that the adversary has a finite monetary budget or a computational bound on the resources to overcome any proposed privacy technology. In other words, the correct real-world model will increasingly appear to be a global passive adversary with some active capabilities. Of course, having no metric provides cold comfort. Syverson also claims that "metrics should not depend on the values of variables for which we cannot make adequate relevant determinations or predictions," such as the number of senders and receivers of a message [4]. Again, needs to be historicized: While it may be true that Tor itself may not know precisely its number of users due to its own technical limits, a global passive adversary such as the NSA can certainly determine the numbers of senders and receivers of a message in Tor, as well as the number of messages sent: Just because you can't count the number of users of a system doesn't mean that the adversary cannot! Furthermore, making the best attempt at accurate models of users and other variables should be part of an experimental design for any research.

However, what Syverson reveals is that his main problem with entropy is that while mixnet designs can be measured in terms of entropy, Tor cannot. So Syverson's real gripe is with the threat model of the Global Passive Adversary (GPA) that can monitor all packets in a network. Syverson wants to uphold Tor as a good enough technical solution against realistic adversaries and so claims that adversaries such as the GPA are inherently "unrealistic," although this was shown to be false: The NSA and possibly other commercial companies are quite close approximations of a GPA due to their control over myriad Internet Exchange Points (IXPs). He then claims local roaming adversaries are more realistic even although they are weaker, even though these adversaries have been shown to be measurable within an entropy-based framework [30]. Yet research shows that adversaries with only a local view of the network can indeed be measured using entropy, as well as active attackers that are actively compromising nodes (as it increases the prior knowledge of adversaries and so reduces entropy) [30]. Even attacks such as the active interference or altering of traffic distributions is clearly amendable to a treatment in terms of entropy, as the difference between any two distributions can be measured using using entropy-based measures like KL divergence.

Although it is inarguable that onion-routing cannot defeat a GPA, due to their "wide distribution" Tor is more "relatively resistant to other kinds of adversaries," although what precise adversaries these are if left unexplained [4]. The key to the relatively strange threat model of Tor is that the adversaries it works well against are those that are nation-states that do not have vast surveillance capacities, at least outside their own (virtual) borders. This makes sense, as Tor is quite useful as a censorship-resistance tool against countries such as Iran, Russia, and (to a lesser extent) China, as they lack the mass surveillance capabilities of the United States. Still, the NSA can likely approximate a GPA. So the weak threat model of Tor is perfectly sufficient against many countries, but seems less plausible when faced with the "Five Eyes" alliance. Of course, this makes sense as the US government would likely not support a privacy-enhancing technology that could defeat its own surveillance capacities. Thus, Syverson claims due to issues with usability and incentives, in practice mixnets and DC-nets would "not scale enough" to protect against the powerful adversaries. Yet current research on DC-nets like Dissent demonstrates the possibility of scaling [31], and real-world deployed mixnets like Nym appear to be able to scale to as many users as Tor [5]. The primary issue facing critics of entropy and the entire information-theoretic paradigm for measuring anonymity is that "despite their critiques of entropy, we do not know definitively what to put in its place" [4]. However, what is arguably worse than having too many metrics is having no metric whatsoever to measure the anonymity provided by a system. Many of the critics of entropy come from an incredibly narrow view of entropy, restricting its usage only to measuring the anonymity of the senders and receivers of a message probabilistically as is done in anonymous communication networks. Yet this weakness may be a strength, as the narrow usage of entropy would allow it to be tested against AI techniques for de-anonymization easily. Furthermore, as then privacy and AI can be unified via the study of entropy, entropy ends up being a metric that most other metrics can be reduced to. Entropy characterizes the anonymity properties of PETs, but also the ability of AI systems to de-anonymize various proposed PETs.

V. Conclusion

It is reasonable to expect an increasing use of AI to empower the collective intelligence of humanity and machines, allowing us to solve a vast variety of problems that have so far remained beyond our cognitive reach. Yet this requires restricting the data it can operate on, and we hypothesize this requires a new kind of AI (or an anti-AI based as PETs) that determines the precise amount of noise needed to prevent malicious AI from harvesting data. Privacy-enhancing technologies such as the proposed noise-generating networks should allow data to become essentially invisible to AI surveillance. Thus, rather than ban or regulate AI, we can reverse the process by which AI works using its own principles. If a malicious AI finds signals in noise, then an anti-surveillance AI can add noise to the existing signal, making it more difficult for patterns to be discovered and so polluting the "training" data available to AI. Training these networks would in principle be similar to generative adversarial networks, but rather than having a neural network be trained to minimize entropy in an adversarial manner, the goal would be to generate enough noise to prevent pattern recognition by a neural network. The key metric would be maximizing the relative entropy.

Taking entropy seriously as a metric would also allow fertile interdisciplinary conversations. Despite the origin of the notion of entropy in thermodynamics, Shannon claimed

von Neumann told him that "no one really knows what entropy really is, so in a debate you will always have the advantage" [32]. Despite–or because of–this fundamental uncertainty about the meaning of entropy, entropy as a measure of noise has been increasingly used to study phenomena ranging from social media [33] to biology [34]. As the massive energy expenditure of generative AI training drives increased interest in sustainable computing systems, entropy could be a valuable concept to help develop sustainable systems; and it should not be forgotten that privacy-enhancing technologies that add noise via cryptography and cover traffic inherently consume more resources than systems without privacy. The use of a thermodynamical conception of entropy in ecological economics has been recently revived, and has even led to the exploration of combining a thermodynamical reading of entropy with the information-theoretic reading of the term by philosophers [35]. Entropy may not only help unify PETs with AI, but help unify computing with the rest of the sciences to tackle current concerns around catastrophic climate change.

Note that we are refraining from discussing the privacy goals of particular AI applications, although adding privacy in some form to AI (for example, to create forms of anonymous AI that does not track its users) is a laudable goal. This is because these particular goals are rather varied. What we are focused on is the abstract paradigm of entropy and how it binds together AI and privacy. However, there is much work to be done. We have only had the time to qualitatively sketch the possible unification of AI and PETs via information theory. A thorough systematization of knowledge of the various uses of entropy across AI and PETs would be the first step. Then the most important next step would be the formalization of the notion of a noise-generating network and if possible, a proof–or proof of non-existence!–of the universal obfuscation theorem could be done. Despite these limitations, a unified conceptual framework for thinking AI and privacy together via the strong foundations of information theory offers tantalizing prospects. As for the future of privacy in the age of AI, wherever there is signal, there can also be noise.

References

[1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems*, 2017.

[2] G. Sebastian, "Privacy and data protection in ChatGPT and other AI Chatbots: strategies for securing user information," *International Journal of Security and Privacy in Pervasive Computing (IJSPPC)*, vol. 15, no. 1, pp. 1–14, 2023.

[3] D. J. Solove, "The virtues of knowing less: Justifying privacy protections against disclosure," *Duke Law Journal*, vol. 53, p. 967, 2003.

[4] P. Syverson, "Why I'm not an entropist," in *International Workshop on Security Protocols*. Springer, 2009, pp. 213–230.

[5] C. Diaz, H. Halpin, and A. Kiayias, "The Nym Network," 2021, https://nym.com/nym-whitepaper.pdf.

[6] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural networks*, vol. 2, no. 5, pp. 359–366, 1989.

[7] B. C. Smith, *The promise of artificial intelligence: reckoning and judgment*. MIT Press, 2019.

[8] T. Maszczyk and W. Duch, "Comparison of Shannon, Renyi and Tsallis entropy used in decision trees," in *Artificial Intelligence and Soft Computing*. Springer, 2008, pp. 643–651.

[9] T. Jaakkola, M. Meila, and T. Jebara, "Maximum entropy discrimination," *Advances in Neural Information Processing Systems*, vol. 12, 1999.

[10] D. J. MacKay, *Information theory, inference and learning algorithms*. Cambridge University Press, 2003.

[11] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Mathematics of control, signals and systems*, vol. 2, no. 4, pp. 303–314, 1989.

[12] V. E. Ismailov, "A three layer neural network can represent any multivariate function," *Journal of Mathematical Analysis and Applications*, vol. 523, no. 1, p. 127096, 2023.

[13] L. Zhou, S. Pan, J. Wang, and A. V. Vasilakos, "Machine learning on big data: Opportunities and challenges," *Neurocomputing*, vol. 237, pp. 350–361, 2017.

[14] Z. Zhang and M. Sabuncu, "Generalized cross entropy loss for training deep neural networks with noisy labels," *Advances in Neural Information Processing Systems*, vol. 31, 2018.

[15] C. E. Shannon, "Prediction and entropy of printed english," *Bell system technical journal*, vol. 30, no. 1, pp. 50–64, 1951.

[16] "ChatGPT and OpenAI models: A preliminary review, author=Roumeliotis, Konstantinos I and Tselikas, Nikolaos D, journal=Future Internet, volume=15, number=6, pages=192, year=2023, publisher=MDPI."

[17] T. D. Vu, H.-J. Yang, V. Q. Nguyen, A.-R. Oh, and M.-S. Kim, "Multimodal learning using convolution neural network and sparse autoencoder," in *IEEE international conference on big data and smart computing (BigComp)*. IEEE, 2017, pp. 309–312.

[18] C. E. Shannon, "A mathematical theory of communication," *The Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.

[19] ——, "A mathematical theory of cryptography," *Mathematical Theory of Cryptography*, 1945.

[20] Y. Dodis and A. Smith, "Entropic security and the encryption of high entropy messages," in *Theory of Cryptography Conference*. Springer, 2005, pp. 556–577.

[21] R. J. McEliece, "A public-key cryptosystem based on algebraic fields," *DSN Progress report*, no. 1978, pp. 114–116, 1978.

[22] A. Serjantov and G. Danezis, "Towards an information theoretic metric for anonymity," in *International Workshop on Privacy Enhancing Technologies*. Springer, 2002, pp. 41–53.

[23] A. R. Beresford and F. Stajano, "Location privacy in pervasive computing," *IEEE Pervasive Computing*, vol. 2, no. 1, pp. 46–55, 2003.

[24] I. Wagner and D. Eckhoff, "Technical privacy metrics: a systematic survey," *ACM Computing Surveys (Csur)*, vol. 51, no. 3, pp. 1–38, 2018.

[25] Y.-A. De Montjoye, C. A. Hidalgo, M. Verleysen, and V. D. Blondel, "Unique in the crowd: The privacy bounds of human mobility," *Scientific reports*, vol. 3, no. 1, pp. 1–5, 2013.

[26] C. Dwork, "Differential privacy," in *International colloquium on automata, langufRages, and programming*. Springer, 2006, pp. 1–12.

[27] N. Li, T. Li, and S. Venkatasubramanian, "t-closeness: Privacy beyond k-anonymity and l-diversity," in *IEEE international conference on data engineering*. IEEE, 2006, pp. 106–115.

[28] D. Lazar, Y. Gilad, and N. Zeldovich, "Karaoke: Distributed private messaging immune to passive traffic analysis," in *USENIX Symposium on Operating Systems Design and Implementation (OSDI)*, 2018, pp. 711–725.

[29] R. Dingledine, N. Mathewson, and P. Syverson, "Tor: The second-generation onion router." in *USENIX Security Symposium*, vol. 4, 2004, pp. 303–320.

[30] I. B. Guirat, C. Diaz, K. Eldefrawy, and H. Zeilberger, "Traffic analysis by adversaries with partial visibility," in *European Symposium on Research in Computer Security*. Springer, 2023, pp. 338–358.

[31] D. I. Wolinsky, H. Corrigan-Gibbs, B. Ford, and A. Johnson, "Dissent in numbers: Making strong anonymity scale," in *USENIX Symposium on Operating Systems Design and Implementation (OSDI)*, 2012, pp. 179–182.

[32] M. Tribus and E. C. McIrvine, "Energy and information," *Scientific American*, vol. 225, no. 3, pp. 179–190, 1971.

[33] H. Halpin, V. Robu, and H. Shepherd, "The complex dynamics of collaborative tagging," in *Proceedings of the 16th international conference on World Wide Web*, 2007, pp. 211–220.

[34] F. Bailly and G. Longo, "Biological organization and anti-entropy," *Journal of Biological Systems*, vol. 17, no. 01, pp. 63–96, 2009.

[35] B. Stiegler, "The ordeal of truth: Causes and quasi-causes in the entropocene," *Foundations of Science*, vol. 27, no. 1, pp. 271–280, 2022.