Project 1: White Paper

CS-370

Brandon Hobbs

March 24, 2023

Neural networks are a type of machine learning (ML) algorithm that attempt to simulate the workings of the human brain. They are used for a wide range of applications, including image recognition, natural language processing, and voice recognition. They are a subset of intelligent systems, from Sheldon (2022), "An intelligent system (1) operates in an environment with other agents, (2) possesses cognitive abilities such as perception, action control, deliberative reasoning or language use, (3) follows behavioral principles based on rationality and social norms and (4) has the capacity to adapt through learning. "

A neural network consists of layers of interconnected nodes, also known as neurons. These artificial neurons according to Gulli and Pal (2017), were inspired by biological processes. Biological (BNN) and artificial (ANN) neurons both seek to take in input and "[output] some kind of signal if the sum of these weighted inputs reach a certain bias" (Nagyfi, 2018).

The first layer, called the input layer, receives the data that the network will be trained on. The data is then passed through one or more hidden layers, where it is processed and transformed. Finally, the output layer produces the network's predictions or classifications based on the input data.

Each neuron in a neural network receives inputs from multiple other neurons, and applies a mathematical function to the sum of those inputs to produce an output. The strength of the connections between neurons, called weights, are adjusted during training in order to optimize the network's performance.

In the case of image recognition, for example, the input layer might receive a pixelated image. The hidden layers would then process this image, identifying patterns and features such as edges and shapes. The output layer would then produce a prediction about what the image represents, such as a cat or a dog.

Overall, neural networks learn to classify objects by iteratively adjusting the connections between neurons in response to feedback about how well their predictions match the correct answers.

Neural networks can be utilized to aid in the personalization of the user experience in several ways. For example:

1. Recommender systems: Neural networks can be trained on a user's past behaviors and preferences to generate personalized recommendations. For example, “YouTube’s recommendations are powered by deep neural networks, which have been trained on human preferences to predict what videos people will like. These Deep Neural Networks …can process enormous amounts of data with a single connection and then make predictions based on these connections” (ReelInReel, n.d.).
2. Natural language processing (NLP): Neural networks can be used to understand and analyze a user's natural language input, such as text or voice commands.
3. Image and video recognition: Neural networks can analyze user-generated content, such as photos and videos, to identify patterns and generate personalized recommendations.
4. Personalized search: Neural networks can be used to personalize search results which allows for greater relevance and targeted search results. Potentially improving user experience.

Neural networks can be used to analyze massive amounts of data and generate personalized recommendations or interactions based on a user's specific needs and preferences.

This personalization can raise several ethical concerns, particularly regarding the potential for hidden biases especially where the algorithms are unknown to the user, i.e., black box. Some significant ethical concerns may arise. For example:

1. Transparency: Neural networks are often considered "black boxes" because they can be difficult to interpret or understand. This lack of transparency can lead to ethical concerns around the potential for hidden biases in the system. Users may not know how the system is making recommendations or decisions, and this lack of transparency can make it challenging to assess or address any potential biases.
2. Data privacy: Neural networks rely on large amounts of data to train and personalize their recommendations. However, there are ethical concerns around the privacy of this data. Personal information (PII) can be collected, analyzed, and used without a user's explicit consent or knowledge, leading to concerns around data privacy and potential misuse of personal information.   
   For example, Uber’s routing can expose information about someone without their consent (BBC News, 2017).
3. Amplification of biases: Neural networks can perpetuate and even amplify existing biases in the data they are trained on. If the training data contains biases (such as gender or racial bias), the neural network may learn and perpetuate those biases in its recommendations, potentially leading to discriminatory or unfair outcomes.   
   For example, facial recognition isn’t flawless. Najibi (2020) shows that across 5 different systems females with darker complexion have the lowest accuracy. Darker males are also less accurate than lighter males. These factors could cause more false positives leading to false arrests and detainments.
4. Lack of user control: In a neural network personalization system, users may have limited control over the recommendations they receive. This lack of control can be concerning from an ethical perspective, as users may be subject to recommendations or decisions that they don't agree with or that have unintended consequences.

Because it's crucial to address these ethical issues, designing systems that are explainable, respect user privacy, monitor for potential biases, and ensure that users have sufficient control over their experiences. To achieve these goals, industry has created answers for these necessary design elements. Responsible Research and Innovation (RRI), for example, “...is a term used by the European Union's Framework Programmes to describe scientific research and technological development processes that take into account effects and potential impacts on the environment and society". Their goal appears to hold research to a higher ethical standard.

Industry has adopted the idea of “right to be forgotten” which is very similar to the ideas of General Data Protection Regulation (GDPR) which seeks to codify this "right". It is a movement that calls for companies to allow people to control their data, especially PII. If a person doesn't want their data included or believes it is being stored unnecessarily, they may request it to be deleted by the company.

The GDPR is a regulation in the European Union that governs the collection, processing, and storage of personal data. Several portions of the GDPR affect personalization, including:

1. Transparency: GDPR requires that organizations provide individuals with clear and concise information about the personal data that is being collected, processed, and stored. This includes information about how the data will be used for personalization purposes.
2. Purpose limitation: GDPR requires personal data be collected and processed for specified, explicit, and legitimate purposes. Personalization must be a legitimate purpose for collecting and processing data, and organizations must ensure that they are not collecting more data than is necessary for personalization.
3. Data minimization: GDPR requires personal data be adequate, relevant, and limited to what is necessary for the purposes for which it is processed. Organizations must ensure that they are not collecting more data than is necessary for personalization, and that the data they collect is accurate and up-to-date.
4. Storage limitation: GDPR requires that personal data be kept in a form that permits identification of individuals for no longer than is necessary for the purposes for which the data is processed. Organizations must ensure that they are not storing personal data for longer than is necessary for personalization purposes.
5. Confidentiality: GDPR requires that personal data be processed in a manner that ensures appropriate security, including protection against unauthorized or unlawful processing, accidental loss, destruction, or damage. Organizations must ensure that personal data used for personalization purposes is kept confidential and secure.
6. Accountability: GDPR requires that organizations be able to demonstrate compliance with the GDPR's principles and obligations. This includes maintaining records of personal data processing activities, conducting data protection impact assessments, and implementing appropriate technical and organizational measures to ensure the security of personal data used for personalization purposes.

In our company, GDPR brings several legal concerns. These concerns may include the following:

1. Data protection and privacy: The use of personal data to train the neural network and personalize the user experience may raise concerns around data protection and privacy. We must comply with relevant data protection laws, such as the GDPR, to ensure that the collection, processing, and storage of personal data is lawful and transparent.  
     
   For compliance purposes our internal auditing team should become aware of, trained for, and being measuring our GDPR compliance for any of our business teams operating in the European Bloc.
2. Discrimination and bias: Neural networks can be prone to perpetuating biases and discrimination. We need to make sure our company has a diverse team helping to train, test, and validate our models before being released to the public.
3. Intellectual property: The use of neural networks may raise questions around intellectual property, particularly if the neural network is trained on copyrighted or proprietary data. Legal should be involved in the vetting of any new data set.
4. Liability: Companies may be held liable for any harm caused by the use of the neural network. This liability is another reason to have our company’s auditing team begin to monitor our usage and creation of ANN technologies.

The auditing team must ensure that we comply with the GDPR's principles of: transparency, purpose limitation, data minimization, accuracy, storage limitation, confidentiality, and accountability. This includes obtaining consent from users for the use of their personal data, ensuring that personal data is accurate and up-to-date, and implementing appropriate technical and organizational measures to protect personal data. If we fail to comply with the GDPR the company may face significant fines and legal liability.

Furthermore, the auditing team should be evaluating the engineering teams’ adherence to industry best practices. Including:

1. Differential privacy: Differential privacy is a machine learning (ML) model technique that adds noise to data to prevent the identification of individual users, enhancing privacy, while still allowing for accurate statistical analysis.
2. Federated learning: Federated learning is a decentralized approach to ML that allows data to be trained locally on user devices, rather than being centralized in one location – helpful in mitigating the damage of data breaches.
3. Homomorphic encryption: Homomorphic encryption is a technique that allows computations to be performed on encrypted data without decrypting it. This technique can be used in ML models to protect user privacy while still allowing for accurate analysis of the data.
4. Privacy-preserving data sharing: Privacy-preserving data sharing techniques allow multiple parties to share data while still maintaining the privacy of individual users.

Our company must be committed to labelling data sensitivity, especially where PII is concerned. We must use encrypted storage anytime PII is available in a dataset and then control access. Data must also be capable of being removed from our systems once we are no longer actively using the data for legitimate purposes – archiving for future use is not acceptable. All of these must be routinely audited for with any failures documented for future resolution.

Overall, the trend in AI and ML is towards a greater focus on privacy and ethics. Companies are increasingly recognizing the importance of protecting user privacy and ensuring that their models are transparent and fair.

Citations:

BBC News. (2017, February 12). *Cheating Frenchman sues Uber for tipping off wife about affair*. https://www.bbc.com/news/world-europe-38948281

Gulli, A., & Pal, S. (2017). Deep learning with keras : Get to grips with the basics of Keras to implement fast and efficient deep-learning models. Packt Publishing, Limited.

Imperva. *What is personally identifiable information*. (n.d.). https://www.imperva.com/ learn/data-security/personally-identifiable-information-pii

Nagyfi, R. (2018). The differences between artificial and biological neural networks. Medium. https://towardsdatascience.com/the-differences-between-artificial-and-biological-neural-networks-a8b46db828b7

Najibi, A. (2020). *Racial discrimination in face recognition technology*. Science in the News. https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology

ReelnReel. (n.d.). *YouTube recommendations: how YouTube uses deep neural networks to recommend videos*. https://reelnreel.com/how-youtube-uses-deep-neural-networks

Responsible Research and Innovation (2023, February 28). In Wikipedia. https://en.wikipedia.org/w/index.php?title=Responsible\_Research\_and\_Innovation&oldid=1142069117

Rudin, C., & Radin, J. (2019, November 22). *Why are we using black box models in ai when we don’t need to? a lesson from an explainable AI competition.* Harvard Data Sci. Rev., 1(2). doi: 10.1162/99608f92.5a8a3a3d

Sheldon, Robert. (2022, June). Intelligent system. Whatis.com. https://www.techtarget.com/whatis/definition/intelligent-system

Spillane, J. (2022, December 8*). How GDPR can undermine personalization and user experience*. Business 2 Community. https://www.business2community.com/customer-experience/how-gdpr-can-undermine-personalization-and-user-experience-02108269

Wiggers, K. (2019, December 21.). *AI has a privacy problem, but these techniques could fix it*. VentureBeat. https://venturebeat.com/ai/ai-has-a-privacy-problem-but-these-techniques-could-fix-it