Computational Stylometry: Programs that Know You

John Walbran

University of Minnesota, Morris

April 2024

- Consider the painting 'The Morteratsch Glacier, Upper Engadine Valley, Pontresina, by. Albert Bierstadt, 1895 [Sethi(2016)]

(a)

(b)

Figure: (a) The original painting (b) The same painting as if it were painted by other artists: (from top left) Van Gogh, Munch, Kahlo, Picasso, Matisse and Escher. [Sethi(2016)]

Outline

- Stylometry
- Neural Networks (NNs)
- Convolutional Neural Networks (CNNs)
- Case Study: Chess
- Ethics
- Conclusion

Stylometry: Definition and History

- Stylometry is the study of identifying the authorship of a work based on its style.

Stylometry: Definition and History

- Stylometry is the study of identifying the authorship of a work based on its style.
- Stylometry has been used to help identify the authors of the works of Shakespeare.

Stylometry: Definition and History

- Stylometry is the study of identifying the authorship of a work based on its style.
- Stylometry has been used to help identify the authors of the works of Shakespeare.
- Shakespeare's canon was proven to be written by multiple authors [Wikipedia([n. d.]a)].

Stylometry: Modern Use Cases

- The advent of machine learning allows the application of stylometry to additional forms of media.

Figure: Example of genuine (upper) and forged (lower) signatures [Hafemann et al.(2017)].

Stylometry: Modern Use Cases

- The advent of machine learning allows the application of stylometry to additional forms of media.
- These include images, sounds, and more accurate analysis of text.

Figure: Example of genuine (upper) and forged (lower) signatures [Hafemann et al.(2017)].

Stylometry: Modern Use Cases

- The advent of machine learning allows the application of stylometry to additional forms of media.
- These include images, sounds, and more accurate analysis of text.
- Stylometry has been used to identify whether signatures are forged or genuine [Hafemann et al.(2017)].

Figure: Example of genuine (upper) and forged (lower) signatures [Hafemann et al.(2017)].

Imitation Stylometry

- Identifying an author of text or media can be extended into imitating that author.

Imitation Stylometry

- Identifying an author of text or media can be extended into imitating that author.
- This can be used to imitate artistic style [Sethi(2016)], or to imitate chess playing [McIlroy-Young et al.(2021)].

Neural Networks: Overview

- Machine learning: automatic process of approximating dependencies of many parameters.

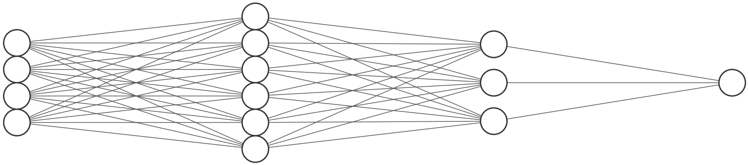


Figure: An overview of a NN structure.

Neural Networks: Overview

- Machine learning: automatic process of approximating dependencies of many parameters.
- Neural Network (NN): a common program model used for machine learning.

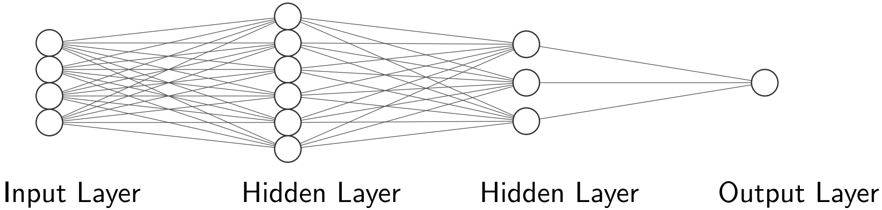


Figure: An overview of a NN structure.

Neural Networks: Overview

- Machine learning: automatic process of approximating dependencies of many parameters.
- Neural Network (NN): a common program model used for machine learning.
- One simple kind of NN is called a Multi-Layered-Perceptrons (MLP).

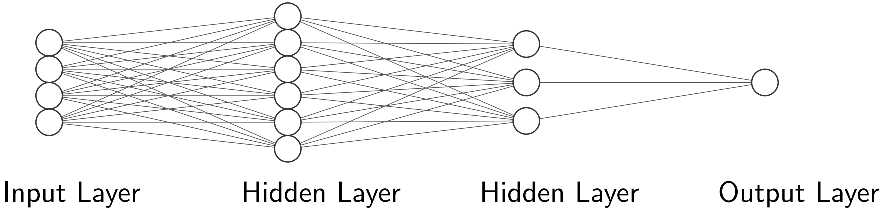


Figure: An overview of a NN structure.

Neural Networks: Overview

- Machine learning: automatic process of approximating dependencies of many parameters.
- Neural Network (NN): a common program model used for machine learning.
- One simple kind of NN is called a Multi-Layered-Perceptrons (MLP).
- NNs consist of connected layers of nodes.

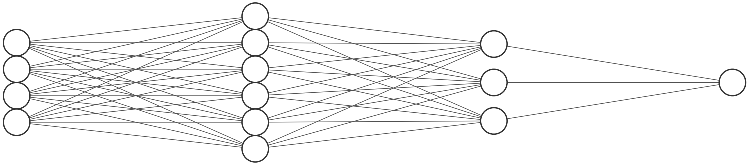
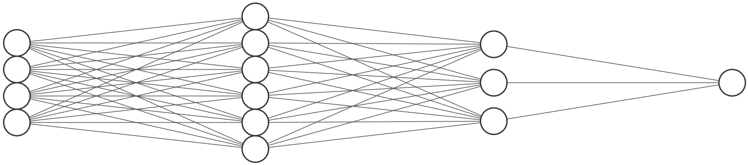


Figure: An overview of a NN structure.

Neural Networks: Overview

- Machine learning: automatic process of approximating dependencies of many parameters.
- Neural Network (NN): a common program model used for machine learning.
- One simple kind of NN is called a Multi-Layered-Perceptrons (MLP).
- NNs consist of connected layers of nodes.
- Each connection has an associated weight.



Input Layer Hidden Layer Hidden Layer Output Layer

Figure: An overview of a NN structure.

Neural Networks: Training and Weights

- Training improves prediction accuracy.

Neural Networks: Training and Weights

- Training improves prediction accuracy.
- This requires data to train the network, with labels indicating correct output.

Neural Networks: Training and Weights

- Training improves prediction accuracy.
- This requires data to train the network, with labels indicating correct output.
- Training starts with random weights.

Neural Networks: Training and Weights

- Training improves prediction accuracy.
- This requires data to train the network, with labels indicating correct output.
- Training starts with random weights.
- Tuning the weights to increase accuracy is an automated process.

Neural Networks: Training and Weights

- Training improves prediction accuracy.
- This requires data to train the network, with labels indicating correct output.
- Training starts with random weights.
- Tuning the weights to increase accuracy is an automated process.
- After training, the weights are frozen, finalizing the model.

Neural Networks: Activation Functions

- Each node of a NN has an associated nonlinear activation function.

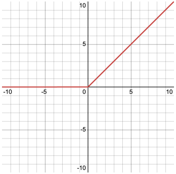


Figure: ReLU(x)

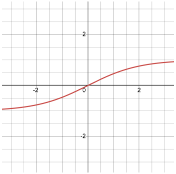


Figure: tanh(x)

- Each node of a NN has an associated nonlinear activation function.
- Examples:
  - ReLU (Rectified Linear Unit):

$$ReLU(x) = \begin{cases} x, x \geq 0 \\ 0 \end{cases}$$

  - tanh:

$$\tanh(x) = \frac{e^{2x} - 1}{e^{2x} + 1}$$

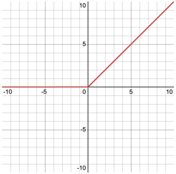


Figure: ReLU(x)

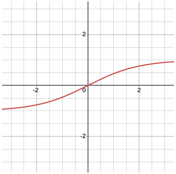


Figure: tanh(x)

Neural Networks: Evaluation

- Evaluation occurs from a layer to each node of the next layer.

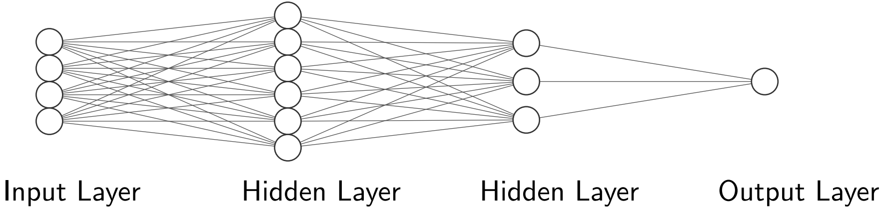


Figure: An overview of a NN structure.

Neural Networks: Evaluation

- Evaluation occurs from a layer to each node of the next layer.
- The result for each node is a linear combination of the previous layer and weights.

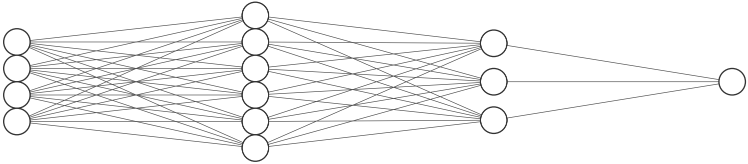


Figure: An overview of a NN structure.

Neural Networks: Evaluation

- Evaluation occurs from a layer to each node of the next layer.
- The result for each node is a linear combination of the previous layer and weights.
- Example: input vector $\vec{x}$, activation function $f(\vec{x})$:

$$N(\vec{x}) = f(x_1 w_1 + x_2 w_2 + \ldots + x_n w_n)$$

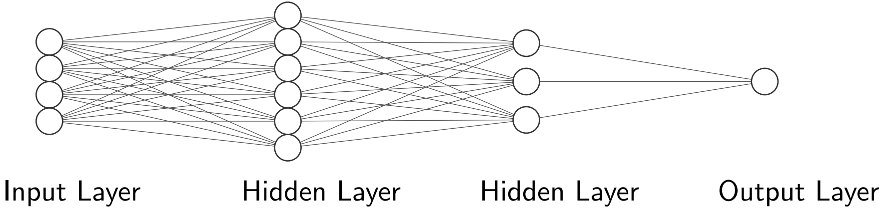


Figure: An overview of a NN structure.

Convolutional Neural Networks

- Convolutional Neural Networks (CNN) are a subset of neural networks that train multi-dimensional data.

Convolutional Neural Networks

- Convolutional Neural Networks (CNN) are a subset of neural networks that train multi-dimensional data.
- They process inputs by preserving the spacial relations between different points.

Convolutional Neural Networks

- Convolutional Neural Networks (CNN) are a subset of neural networks that train multi-dimensional data.
- They process inputs by preserving the spacial relations between different points.
- These often work on 2D images.

Convolutional Neural Networks

- Convolutional Neural Networks (CNN) are a subset of neural networks that train multi-dimensional data.
- They process inputs by preserving the spacial relations between different points.
- These often work on 2D images.
- CNNs can be thought of as a traditional NN with convolutional steps.

CNNs: Convolutions

- A convolutional filter is a grid of weights that get multiplied element-wise with each subset of the input.

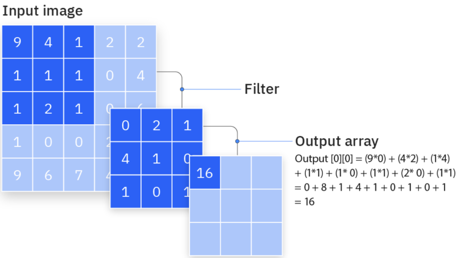


Figure: An example convolutional filter being combined with one subset of the input [IBM([n. d.])].

CNNs: Convolutions

- A convolutional filter is a grid of weights that get multiplied element-wise with each subset of the input.
- The weights of the filter are what gets trained for the CNN.

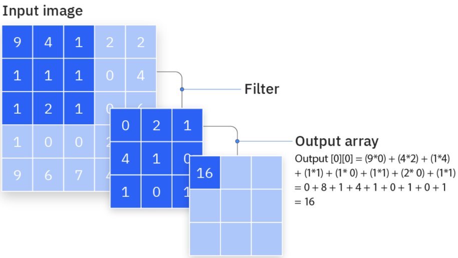


Figure: An example convolutional filter being combined with one subset of the input [IBM([n. d.])].

CNNs: Residual CNNs

- CNNs often train for different features simultaneously.

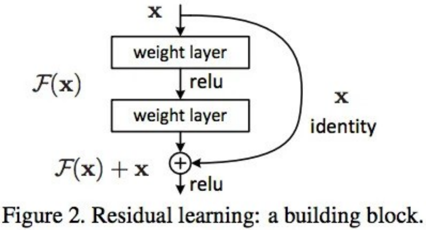


Figure: [Shorten([n. d.])]

CNNs: Residual CNNs

- CNNs often train for different features simultaneously.
- Original details get lost after CNN layers.

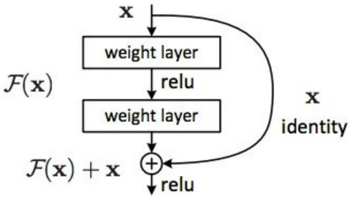


Figure 2. Residual learning: a building block.

Figure: [Shorten([n. d.])]

CNNs: Residual CNNs

- CNNs often train for different features simultaneously.
- Original details get lost after CNN layers.
- Residual CNNs pass original data with isolated features to preserve fidelity.

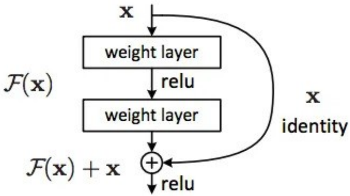


Figure 2. Residual learning: a building block.

Figure: [Shorten([n. d.])]

Case Study: Chess

- McIlroy-Young *et al.* created a stylometric model to identify chess players by their games [McIlroy-Young et al.(2021)].

Figure: [Wikipedia([n. d.]b)]

Case Study: Chess

- McIlroy-Young *et al.* created a stylometric model to identify chess players by their games [McIlroy-Young et al.(2021)].
- They did this in the hopes of building personalized training assistants [McIlroy-Young et al.(2021)].

Figure: [Wikipedia([n. d.]b)]

Case Study: Chess

- McIlroy-Young *et al.* created a stylometric model to identify chess players by their games [McIlroy-Young et al.(2021)].
- They did this in the hopes of building personalized training assistants [McIlroy-Young et al.(2021)].
- They described their process and results over a series of papers, with the most recent being *Learning Models of Individual Behavior in Chess*, published in 2022.

Figure: [Wikipedia([n. d.]b)]

Chess: Training Data

- Mcillroy-Young *et al.* used data from lichess.org

Chess: Training Data

- Mcillroy-Young *et al.* used data from lichess.org
- The lichess database has more than a billion games growing by more than 1 million games a day [McIlroy-Young et al.(2021)].

Chess: Training Data

- Mcillroy-Young *et al.* used data from lichess.org
- The lichess database has more than a billion games growing by more than 1 million games a day [McIlroy-Young et al.(2021)].
- Game data contains metadata, and all moves.
    - Player identifiers
    - Player ratings
    - Time control
    - Event (if applicable)

Chess Engine: Training Setup

- Players were filtered by:
    - Active in December 2020.
    - Has played more than 1000 blitz games.
    - Has a mean rating between 1000-2000.
    - Has low rating variance.

Chess Engine: Training Setup

- Players were filtered by:
  - Active in December 2020.
  - Has played more than 1000 blitz games.
  - Has a mean rating between 1000-2000.
  - Has low rating variance.
- Players were further grouped into categories based on total number of games played.

Chess Engine: Training Setup

- Players were filtered by:
  - Active in December 2020.
  - Has played more than 1000 blitz games.
  - Has a mean rating between 1000-2000.
  - Has low rating variance.
- Players were further grouped into categories based on total number of games played.
- Games with $<$10 moves were discarded.

Chess Engine: Training Setup

- Players were filtered by:
  - Active in December 2020.
  - Has played more than 1000 blitz games.
  - Has a mean rating between 1000-2000.
  - Has low rating variance.
- Players were further grouped into categories based on total number of games played.
- Games with $<$10 moves were discarded.
- Games were split into three categories for training:
  - Training games.
  - Reference games.
  - Query games.

Chess Engine: Training Setup

- Players were filtered by:
  - Active in December 2020.
  - Has played more than 1000 blitz games.
  - Has a mean rating between 1000-2000.
  - Has low rating variance.
- Players were further grouped into categories based on total number of games played.
- Games with $<$10 moves were discarded.
- Games were split into three categories for training:
  - Training games.
  - Reference games.
  - Query games.
- Only considered 100 reference and query games each.

Chess Engine: Architecture

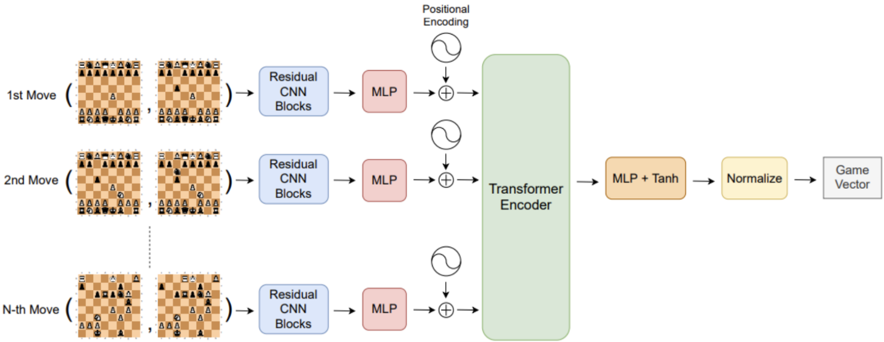


Figure: The architecture of the neural network used by Mcillroy-Young *et al.* [McIlroy-Young et al.(2021)].

Chess Engine: Input

- The model takes in a sequence of moves.

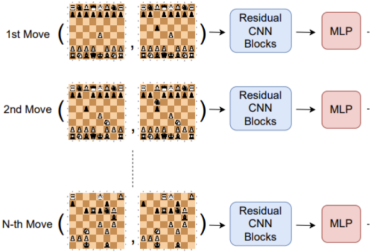


Figure: Model architecture for input handling [McIlroy-Young et al.(2021)].

Chess Engine: Input

- The model takes in a sequence of moves.
- Moves are represented as positions before-and-after.

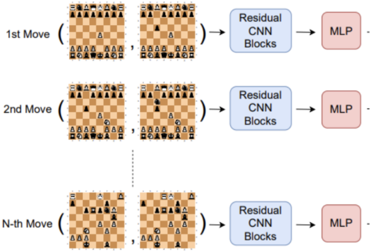


Figure: Model architecture for input handling [McIlroy-Young et al.(2021)].

Chess Engine: Input

- The model takes in a sequence of moves.
- Moves are represented as positions before-and-after.
- Positions are 2D boards for each piece type and metadata.

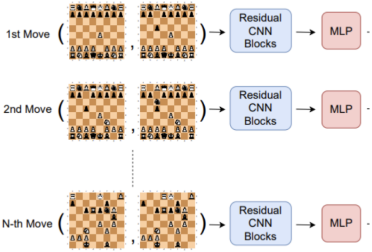


Figure: Model architecture for input handling [McIlroy-Young et al.(2021)].

Chess Engine: Input

- The model takes in a sequence of moves.
- Moves are represented as positions before-and-after.
- Positions are 2D boards for each piece type and metadata.
- These moves are fed into a residual CNN, outputting move features.

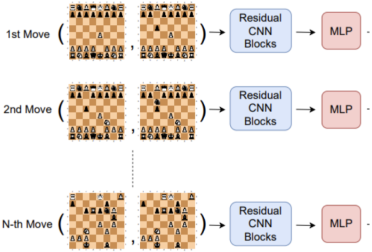


Figure: Model architecture for input handling [McIlroy-Young et al.(2021)].

Chess Engine: Processing and Output

- Move features are passed into transformer.

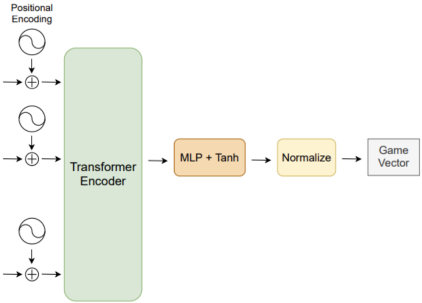


Figure: Model architecture for processing and output.

Chess Engine: Processing and Output

- Move features are passed into transformer.
- The transformer takes all move features from the sequence.

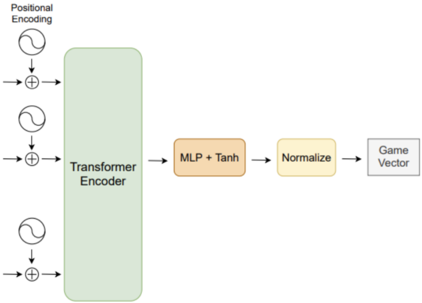


Figure: Model architecture for processing and output.

Chess Engine: Processing and Output

- Move features are passed into transformer.
- The transformer takes all move features from the sequence.
- The move features are then compressed into their essence, creating a game vector.

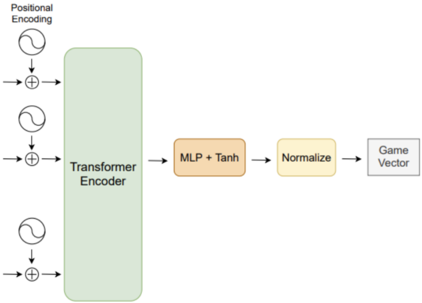


Figure: Model architecture for processing and output.

Chess Engine: Game Vectors

- Game vectors represent the essence of a game.

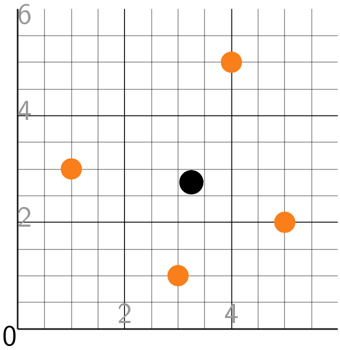


Figure: The centroid of a set of points.

Chess Engine: Game Vectors

- Game vectors represent the essence of a game.
- Taking the centroid of games by a player gives a player identity.

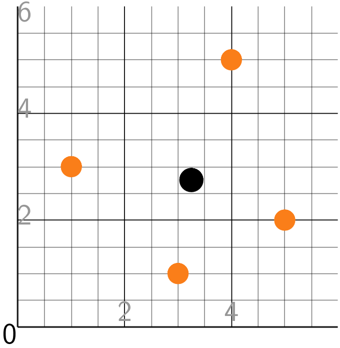


Figure: The centroid of a set of points.

Chess Engine: Game Vectors

- Game vectors represent the essence of a game.
- Taking the centroid of games by a player gives a player identity.
- New players added by same process.

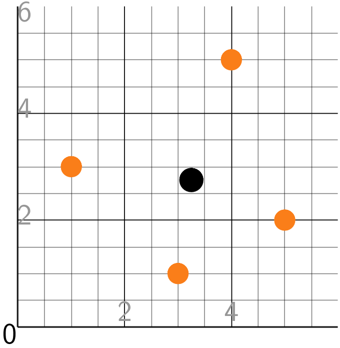


Figure: The centroid of a set of points.

Chess Engine: Results

- Tested with players with more than 10K games.

All data from [McIlroy-Young et al.(2021)].

Chess Engine: Results

- Tested with players with more than 10K games.
- Their model correctly identified players 86% of the time.

All data from [McIlroy-Young et al.(2021)].

Chess Engine: Results

- Tested with players with more than 10K games.
- Their model correctly identified players 86% of the time.
- This increased to 92% when considering similar rating.

All data from [McIlroy-Young et al.(2021)].

Chess Engine: Results

- Tested with players with more than 10K games.
- Their model correctly identified players 86% of the time.
- This increased to 92% when considering similar rating.
- Accuracy for unseen players was 85%.

All data from [McIlroy-Young et al.(2021)].

Chess Engine: Results

- Tested with players with more than 10K games.
- Their model correctly identified players 86% of the time.
- This increased to 92% when considering similar rating.
- Accuracy for unseen players was 85%.
- Model generalized to master players.

All data from [McIlroy-Young et al.(2021)].

Ethics

- McIlroy-Young *et al.* were asked to include a companion paper on ethics [Hutson(2022)].

Ethics

- McIlroy-Young *et al.* were asked to include a companion paper on ethics [Hutson(2022)].
- They published *Mimetic Models: Ethical Implications of AI that Acts Like You* with their other work in 2022.

Ethics

- McIlroy-Young *et al.* were asked to include a companion paper on ethics [Hutson(2022)].
- They published *Mimetic Models: Ethical Implications of AI that Acts Like You* with their other work in 2022.
- Mimetic models are imitation models.

Ethics

- McIlroy-Young *et al.* were asked to include a companion paper on ethics [Hutson(2022)].
- They published *Mimetic Models: Ethical Implications of AI that Acts Like You* with their other work in 2022.
- Mimetic models are imitation models.
- They focus on examples of imitation, challenging privacy and value.

Ethics

- McIlroy-Young *et al.* were asked to include a companion paper on ethics [Hutson(2022)].
- They published *Mimetic Models: Ethical Implications of AI that Acts Like You* with their other work in 2022.
- Mimetic models are imitation models.
- They focus on examples of imitation, challenging privacy and value.
- If an imitation model is accurate, value of the human subject is devalued.

Ethics

- McIlroy-Young *et al.* were asked to include a companion paper on ethics [Hutson(2022)].
- They published *Mimetic Models: Ethical Implications of AI that Acts Like You* with their other work in 2022.
- Mimetic models are imitation models.
- They focus on examples of imitation, challenging privacy and value.
- If an imitation model is accurate, value of the human subject is devalued.
- Imitation models can be used as a learning tool.

Ethics

- McIlroy-Young *et al.* were asked to include a companion paper on ethics [Hutson(2022)].
- They published *Mimetic Models: Ethical Implications of AI that Acts Like You* with their other work in 2022.
- Mimetic models are imitation models.
- They focus on examples of imitation, challenging privacy and value.
- If an imitation model is accurate, value of the human subject is devalued.
- Imitation models can be used as a learning tool.
- Imitation stylometry can be used for effective counterfeiting.

Conclusion

- Stylometry is a rapidly growing field.

Conclusion

- Stylometry is a rapidly growing field.
- Stylometry can help tailor ML models for individuals.

Conclusion

- Stylometry is a rapidly growing field.
- Stylometry can help tailor ML models for individuals.
- Powerful stylometry raises many privacy concerns.

Acknowledgments

I would like to thank Prof. Elena Machkasova for advising through this project, and Prof. Wenkai Guan for feedback and suggestions during this project.

Bibliography

Hafemann, Sabourin, and Oliveira. 2017.
Learning Features for Offline Handwritten Signature Verification using Deep Convolutional Neural Networks.
Pattern Recognition (2017).
https://doi.org/10.48550/arXiv.1705.05787

Hutson. 2022.
AI unmasks anonymous chess players, posing privacy risks.
https://www.science.org/content/article/ai-unmasks-anonymous-chess-players-posing-privacy-risks

IBM. [n. d.].
What are convolutional neural networks?
https://www.ibm.com/topics/convolutional-neural-networks

McIlroy-Young, Wang, Sen, Kleinberg, and Anderson. 2021.
Detecting Individual Decision-Making Style: Exploring Behavioral Stylometry in Chess.
NeurIPS 2021 34 (2021), 23 pages.
https://doi.org/10.48550/arXiv.2208.01366

Sethi. 2016.
Using computers to better understand art.
https://theconversation.com/using-computers-to-better-understand-art-56887

Shorten. [n. d.].
https://towardsdatascience.com/introduction-to-resnets-c0a830a288a4

Wikipedia. [n. d.]a.
https://en.wikipedia.org/wiki/Shakespeare_attribution_studies

Wikipedia. [n. d.]b.
https://en.wikipedia.org/wiki/Chess