Welcome to the

CoGrammar

Natural Language Processing for Text Analysis

The session will start shortly...

Questions? Drop them in the chat. We'll have dedicated moderators answering questions.

Cyber Security Session Housekeeping

- The use of disrespectful language is prohibited in the questions, this is a supportive, learning environment for all - please engage accordingly. **(Fundamental British Values: Mutual Respect and Tolerance)**
- No question is daft or silly - **ask them!**
- There will be a **Q&A** at the end of the session, should you wish to ask any follow-up questions.
- If you have any questions outside of this lecture, or that are not answered during this lecture, please do submit these for upcoming Academic Sessions. You can submit these questions here: **Questions**

Cyber Security Session Housekeeping cont.

- For all **non-academic questions**, please submit a query: **www.hyperiondev.com/support**
- We would love your **feedback** on lectures: **Feedback on Lectures**
- Find all the lecture **content** in you **Lecture Backpack** on GitHub.
- If you are **hearing impaired**, please kindly **enable captions** on Google chrome/Microsoft Edge via the accessibility settings.

Safeguarding & Welfare

We are committed to all our students and staff feeling safe and happy; we want to make sure there is always someone you can turn to if you are worried about anything.

If you are feeling upset or unsafe, are worried about a friend, student or family member, or you feel like something isn't right, speak to our safeguarding team:

Ian Wyles
Designated Safeguarding Lead

Simone Botes

Nurhaan Snyman

Rafiq Manan

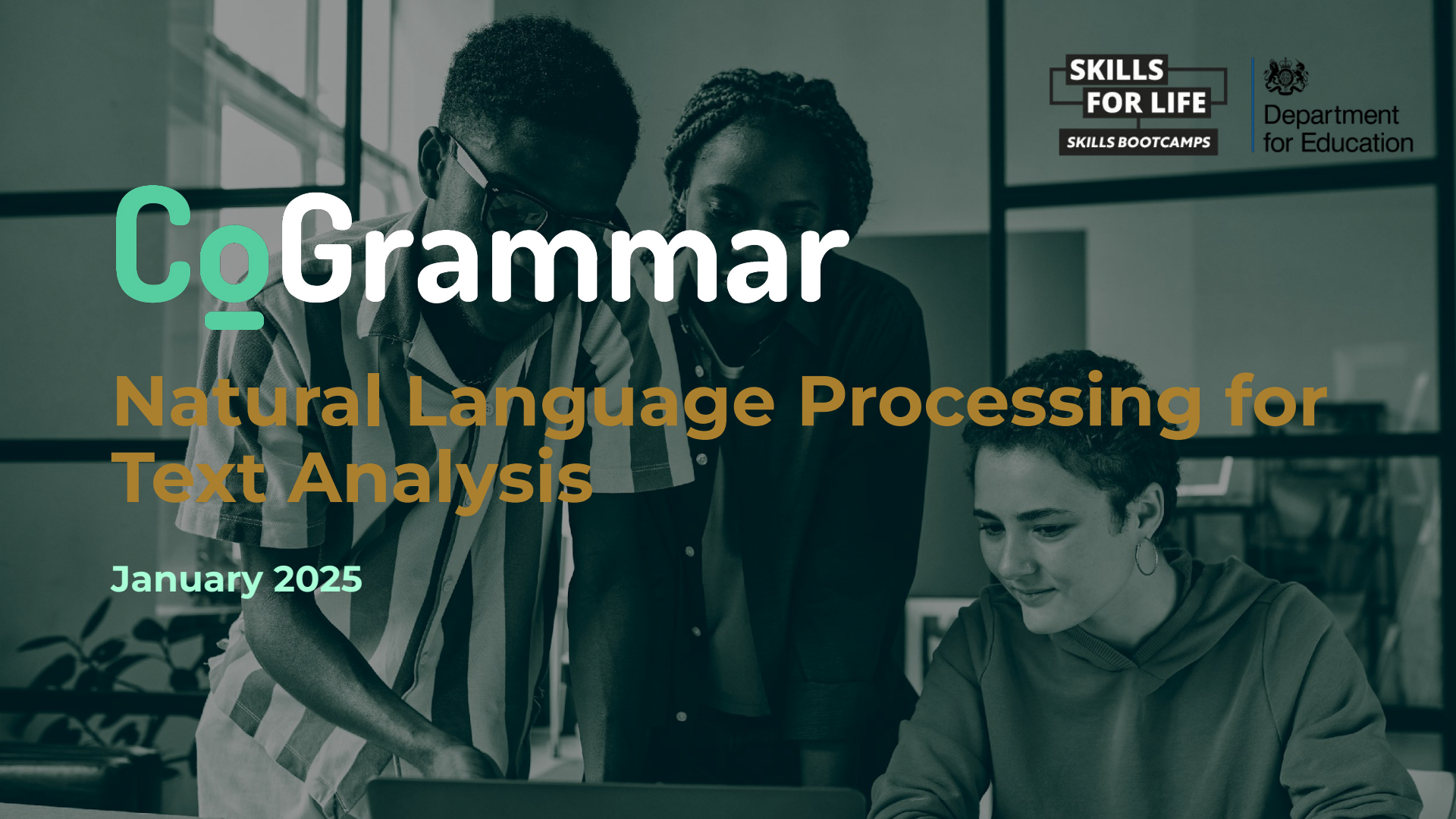
Ronald Munodawafa

Tevin Pitts

Scan to report a safeguarding concern

or email the Designated Safeguarding Lead:
Ian Wyles
safeguarding@hyperiondev.com

CoGrammar HyperionDev



Natural Language Processing for Text Analysis

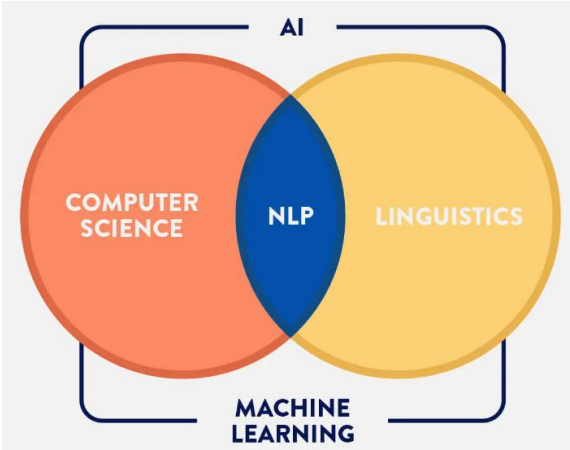
January 2025

Learning Outcomes

1. Understand the basics of Natural Language Processing (NLP) and its applications in cybersecurity.
2. Be familiar with the Jupyter Notebook layout, functionality, and usage.
3. Apply fundamental NLP techniques such as loading and using pre-trained language models in spaCy and utilising extensions like `spacytextblob` for sentiment analysis.
4. Analyse sentiment polarity and interpret its significance in a cybersecurity context.

Natural Language Processing

NLP is a branch of AI that enables computers to comprehend, generate, and manipulate human language.



Source: https://clevertap.com/blog/natural-language-processing/

Natural Languages

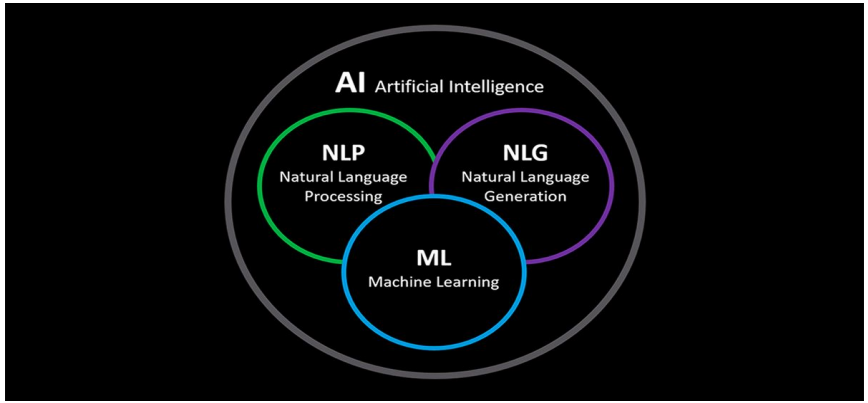
- Natural language = human language (not programming language)
  - **Phonetics and Phonology** - linguistic sounds.
  - **Morphology** - meaningful components of words.
  - **Syntax** - structural relationships between words.
  - **Semantics** - knowledge of meaning.
  - **Pragmatics** - relationship of meaning to goals and intentions of speaker.
  - **Discourse** - linguistic units larger than a single utterance.
- Lots of stuff for a computer to know!

Ambiguity

- Scenario - Tony is telling JARVIS that he was annoyed at a female worker, and threw a piece of paper. He then proceeds to say “I made her duck.”
- How does JARVIS interpret this?
  - I cooked a duck for her.
  - I cooked a duck belonging to her.
  - I created the duck that she owns.
  - I caused her to quickly lower her head.
  - I turned her into a duck (possibly with the help of Dr. Strange?).
- English (and natural languages generally) is weird and ambiguous. This makes it difficult for computers to navigate the way we speak.

How does NLP work?

- Goal is to simulate human intelligence.
- Machine Learning techniques used to train a model to understand human language.
- Probability-based - therefore not always 100% accurate, but close enough, and getting closer every day!



SpaCy

- A software package that comes with all of these models already!
- We won't need to be generating models and training with data.
- We give it a sentence, it gives us linguistic data about the sentence.

The Importance of NLP in Cyber Security

- **Detect phishing emails** by analysing suspicious language and sender details.
- **Automate threat intelligence** from forums, blogs, and social media.
- **Analyse malware scripts** to identify malicious patterns.
- **Monitor online sentiment** for potential threats against organisations.
- **Secure chatbots** by identifying **malicious or exploitative queries.**

Code Demo

CoGrammar

Polls

Please have a look at the poll notification and select an option.

What is the primary purpose of of Natural Language Processing?

A. To process and analyse text and speech data.
B. To secure networks from cyberattacks.
C. To optimize database queries.

Polls

Please have a look at the poll notification and select an option.

What does `doc._.blob.polarity` provide?

A. The subjectivity score of the text.
B. The total word count in the text.
C. The sentiment score of the text.

Polls

Please have a look at the poll notification and select an option.

Which of the following could sentiment analysis be used for in cybersecurity?

A. Detecting phishing emails with negative sentiment.
B. Identifying trends in customer reviews of security products.
C. Monitoring online discussions for potential cyber threats.
D. All of the above.

Questions and Answers

CoGrammar

Thank you for attending

SKILLS FOR LIFE
SKILLS BOOTCAMPS

Department for Education

CoGrammar