AI Toolkit: Libraries and Essays for Exploring the Technology and Ethics of AI

Levin Ho, Morgan McErlean, Zehua (Alex) You, Lisa Meeden
Computer Science Department
Swarthmore College

Douglas Blank
Head of Research, Comet ML

Outline for this talk

1. Why we created the AI toolkit, called AITK
2. What's available in AITK
3. An example from AITK
4. How AITK has been used in courses
5. Conclusion

SWARTHMORE

1. Why we created AITK

AI Literacy

AI's impact on society is at an all time high, and we need to provide ways for our students and the general public to gain a better understanding of AI:

- How does it work?
- What can it do well?
- What are its limitations?
- What ethical issues do we face in its use?

However, most available explanations on how AI works are either too superficial and lack detail or are too complex for novices

SWARTHMORE

Goal: Create a set of accessible AI tools and essays suitable for novices

Collaborative effort built on years of experience of teaching AI at small liberal arts colleges

- Douglas Blank — Bryn Mawr College
- James Marshall — Sarah Lawrence College
- Lisa Meeden — Swarthmore College

AITK was first released as an open source project in 2021

In the summer of 2024, three Swarthmore undergrads added content on Generative AI to AITK

Levin Ho

Morgan McErlean

Zehua (Alex) You

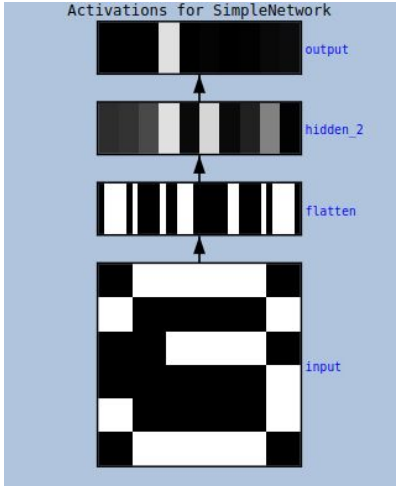
SWARTHMORE

2. What’s Available in AITK

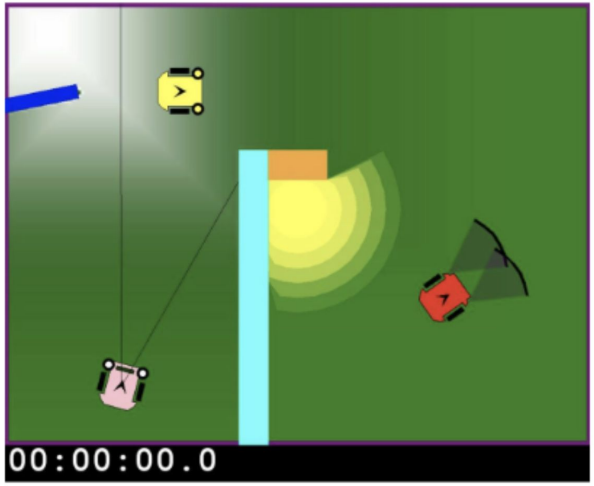
- Python libraries
- Jupyter notebooks

Python libraries

- `aitk.networks`: provides an easy to use wrapper around Keras and Tensorflow with features for better visualizing the inner workings of network models



- `aitk.robots`: provides a robot simulator where multiple robots with different features can interact in a variety of user-defined worlds



SWARTHMORE

Topics and Sequencing of AITK Jupyter Notebooks
Highlighted notebooks focus on societal impacts of AI

	Neural Networks	Generative AI	Robotics
Begin	• Basic Neural Networks	• Word Embedding	• What is it Like to be a Robot?
Next	• Categorizing Faces • Data Manipulation	• Nano GPT • Image Generation	• Braitenberg Vehicles • Seek Light
Further	• Analyzing Hidden Representations • Structure of Convolutional Networks	• Transformer	• Subsumption

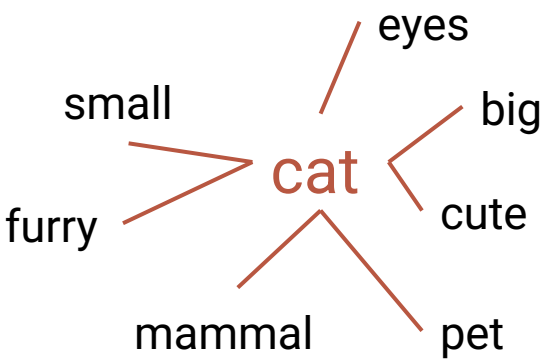
SWARTHMORE

3. Example Notebook from AITK

Example AITK Notebook: Word Embedding

- How can a computer model, with no experience of the real world, gain an understanding of what words mean?
- All neural network models use n-dimensional vectors, called word embeddings, to represent words internally
- These vectors are learned based on how words co-occur with other words in lots of example text

"the *small*, *furry* cat"
..."cats and other *mammals*"
..."the cat's *big eyes*"
..."a *cute pet*, like a cat"



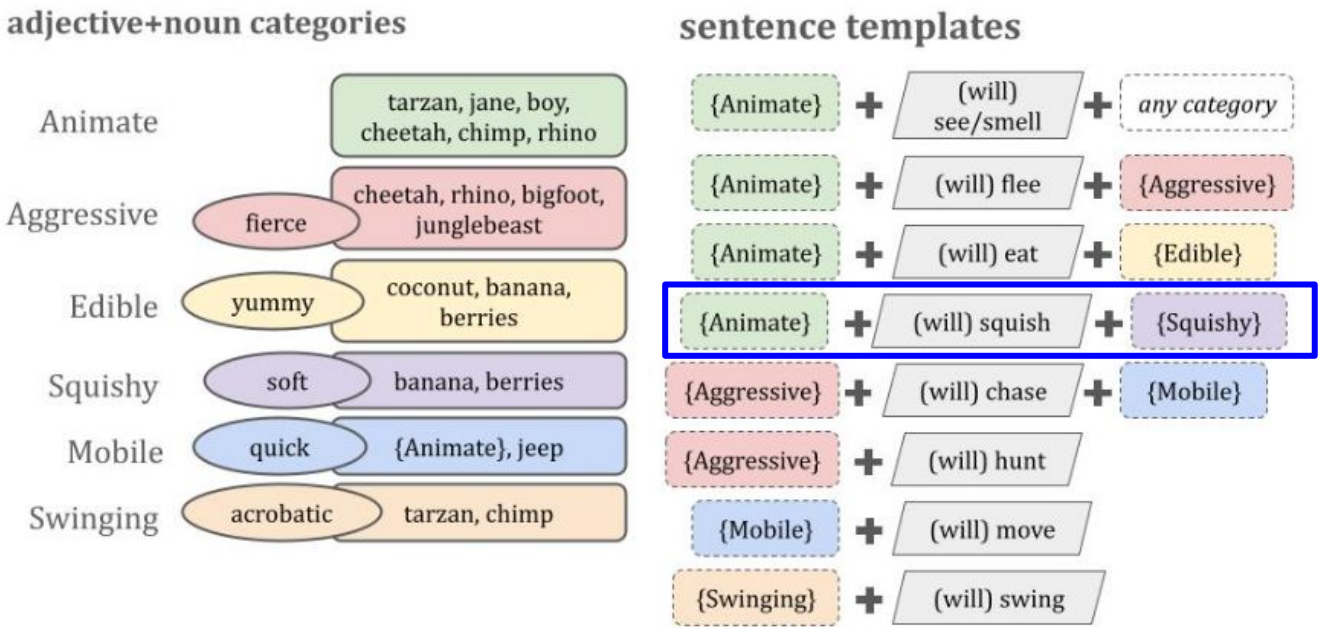
SWARTHMORE

A Simple Word Embedding Model

Tarzan Language

For this notebook, we'll construct a toy language with just 27 words; 13 nouns, 9 verbs, 5 adjectives, and "will". Our nouns and adjectives come in categories, and we'll generate sentences using simple templates based on those categories:

From the Word Embedding notebook



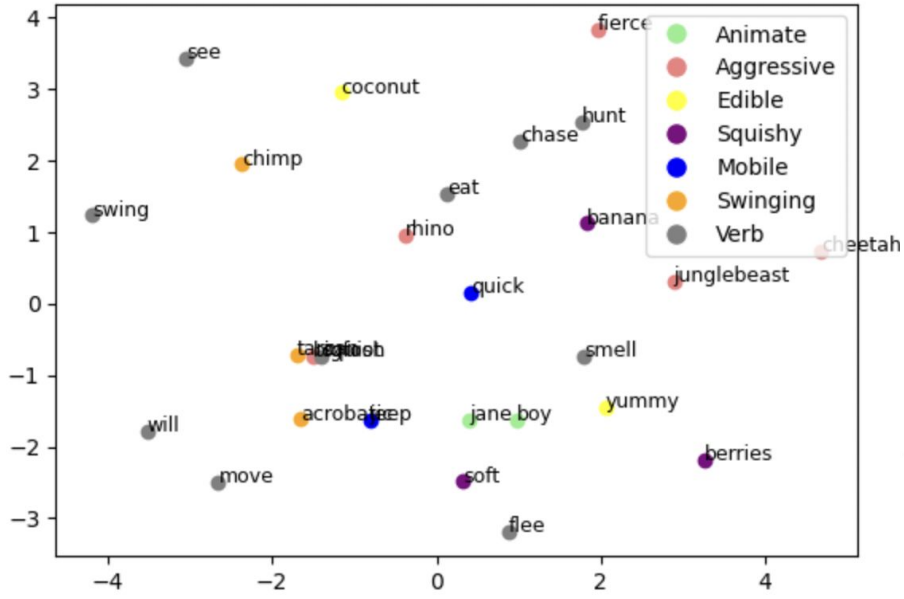
SWARTHMORE

```
Generated 796 sentences. Here are 30 examples:
        chimp see berries
        fierce cheetah chase bigfoot
        boy smell bigfoot
        junglebeast will chase quick rhino
        chimp see cheetah
        chimp flee bigfoot
        cheetah will chase boy
        tarzan see bigfoot
        fierce rhino will chase jeep
        jane will see rhino
        jane will see jeep
        boy flee junglebeast
        chimp smell rhino
        boy will flee bigfoot
        tarzan see boy
        fierce cheetah chase jeep
        rhino will eat berries
        cheetah will see jeep
        jane squish soft banana
        cheetah chase boy
        rhino will eat yummy coconut
        jane squish banana
        jane will squish soft berries
        junglebeast chase quick jane
        junglebeast will chase quick rhino
        jane will squish soft berries
        boy flee cheetah
        chimp will smell banana
        jane will smell cheetah
        fierce bigfoot will hunt
```

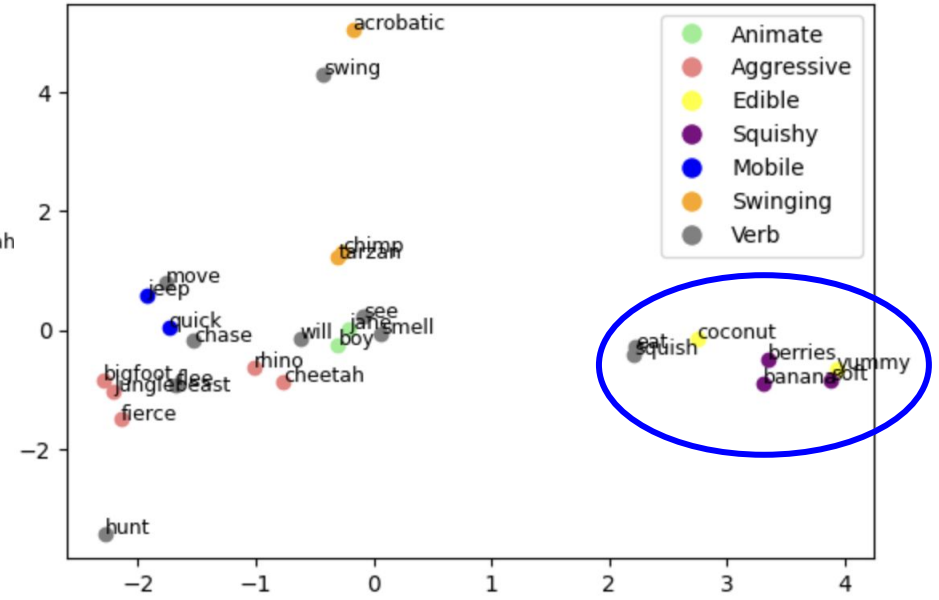
From the Word Embedding notebook

SWARTHMORE

Visualizing Word Embeddings Before and After Training



Before training



After training

SWARTHMORE

Societal Impacts: LLMs May Perpetuate Bias

- Word embedding algorithms do an impressive job of learning to represent the “meanings” of words from associations with other words
- But issues arise when those associations are biased
- The notebook discusses the paper *Man is to Computer Programmer as Woman is to Homemaker?* which shows that embeddings can encode inappropriate gender stereotypes

SWARTHMORE

4. How AITK has been used in various classes

National Humanities Center: Responsible AI Curriculum Design Program 2021 -2024

Institution	Course	Instructor(s)	Semester
Arizona State University	Human Impacts of AI	Gaymon Bennett, Erica O'Neil	Fall 2023
Bowdoin College	AI Ethics	Eric Chown, Allison Cooper, Michael Franz, Fernando Nascimento	Fall 2023
Case Western Reserve University	Responsible AI: Cultivating a Just and Sustainable Socio-technical Future through Data Citizenship	Timothy Beal, Michael Hemenway	Spring 2024
Davidson College	Critical AI Studies	Raghu Ramanujan, Mark Sample	Spring 2024
Duke University	Artificial Intelligence in Literature and Film	Aarthi Vadde (NHC Fellow, 2020–21)	Spring 2024
George Mason University	Equitable AI	Nupoor Ranade	Spring 2024
Johnson C. Smith University	Responsible Artificial Intelligence	Felesia Stukes	Spring 2024
North Carolina State University	Responsible AI and Society	Huiling Ding	Fall 2023
Rice University	Responsible AI for Health	Kirsten Ostherr	Fall 2023
Swarthmore College	Ethics and Technology	Lisa Meeden, Krista K. Thomason (NHC Fellow, 2021–22)	Spring 2024
Texas A&M University	Ethics of Artificial Intelligence	Glen Miller	Fall 2023
University of California, Santa Cruz	Artificial Intelligence and Human Imagination	Zac Zimmer	Fall 2023
University of Florida	Gender, Race, and Worldbuilding with AI	Hina Shaikh	Fall 2023
University of Georgia	AI for Humans: Learning to Live with AI	Kimberly Van Orman	Spring 2024
The University of Utah	Praxis Lab in Responsible AI	Elizabeth Callaway	Fall 2023, Spring 2024

SWARTHMORE

Spring 2024 *Ethics and Technology* with Krista Thomason

Used Notebooks on

- Basic neural networks
- Classifying faces
- What is it like to be a robot?

Comments from students:

“I loved the labs! … I feel they gave me a really good understanding of how ML works.”

“I enjoyed the hands-on nature of the labs.”

“The lab component was very helpful. It was quite lightweight (people without a lot of experience will understand).”

“The lab was super helpful to put our arguments in class into context. “

SWARTHMORE

Spring 2024
AI in Literature and Film @ Duke
Aarthi Vadde

Used Notebook on

- Basic neural networks

Comments from Aarthi Vadde:

Doing this exercise led to a number of interesting discussions in class about:

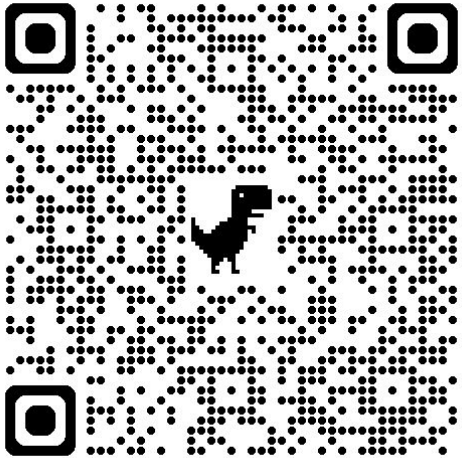
- What experts in specialized fields owe the public
- The degree of literacy the public should have about neural networks and other AI tools

SWARTHMORE

5. Conclusions

Conclusion

- AITK is an open-source project available on github consisting of both Python libraries and Jupyter notebooks
- Designed so that novices can gain a better understanding of the technology and ethics of AI
- AITK has been successfully piloted in many different types of courses at a variety of institutions
- Lightweight and easy to add to an existing course to address AI Literacy



QR code for the Word Embedding AITK Notebook

SWARTHMORE