SIGGRAPH ASIA 2020 VIRTUAL

December 2020
Actual Dates to be advised.
ONLINE EVENT

SA2020.SIGGRAPH.ORG
#SIGGRAPHAsia | #SIGGRAPHAsia2020

Check for updates

eXplainable AI (XAI):

An introduction to the XAI landscape with practical examples Rowan Test

Rowan Hughes[1], Cameron Edmond[1], Lindsay Wells[1], Mashhuda Glencross[2,3], Liming Zhu[3], Tomasz Bednarz[1,3]

[1] UNSW, Sydney, Australia
[2] UQ, Brisbane, Australia
[3] CSIRO's Data61, Australia

Sponsored by

Organized by koelnmesse

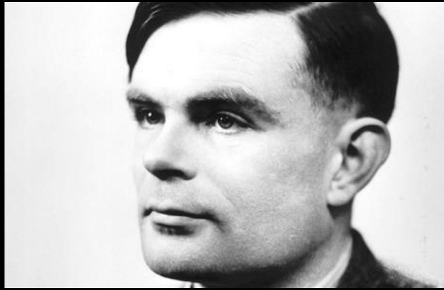
Agenda

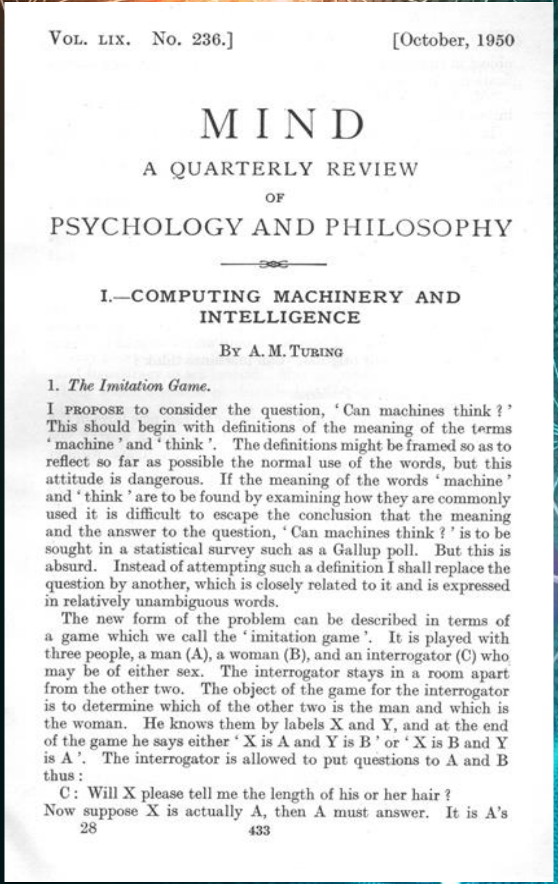
1. **9:00 am**
   - *Introduction*
   - *Agenda*
   - *Background*
2. **9.10 am**
   - *Building Trust in AI*
   - *Responsible Technology*
3. **9.15 am**
   - *XAI Landscape and challenges*
   - *Example 1*
4. **9.40 am**
   - *Generative Networks and XAI*
   - *Example 2*
5. **10:05 am**
   - *Narrative Solutions for XAI*
   - *Narrative Visualization for XAI*
6. **10.25 am**
   - *Wrap Up and Q&A*

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.
Copyright is held by the owner/author(s).
SA '20 Courses, December 04-13, 2020, Virtual Event, Republic of Korea
ACM 978-1-4503-8112-3/20/11.
10.1145/3415263.3419166

Computing and Intelligence

- Do Machine Learning algorithms have a Soul?
- Could they understand every day's reality as us Humans do?
- What the consequence of their Creativity?
- Can they help us to understand world better?





VOL. LIX. No. 236.] [October, 1950

MIND

A QUARTERLY REVIEW OF PSYCHOLOGY AND PHILOSOPHY

I.—COMPUTING MACHINERY AND INTELLIGENCE

BY A. M. TURING

1. *The Imitation Game.*

I PROPOSE to consider the question, 'Can machines think ?' This should begin with definitions of the meaning of the terms 'machine' and 'think'. The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words 'machine' and 'think' are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, 'Can machines think ?' is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words.

The new form of the problem can be described in terms of a game which we call the 'imitation game'. It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and at the end of the game he says either 'X is A and Y is B' or 'X is B and Y is A'. The interrogator is allowed to put questions to A and B thus:

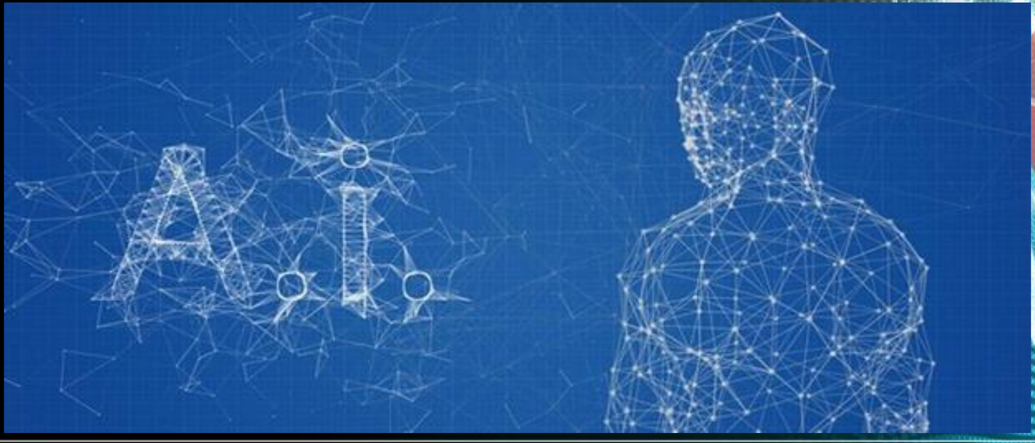
C: Will X please tell me the length of his or her hair ?

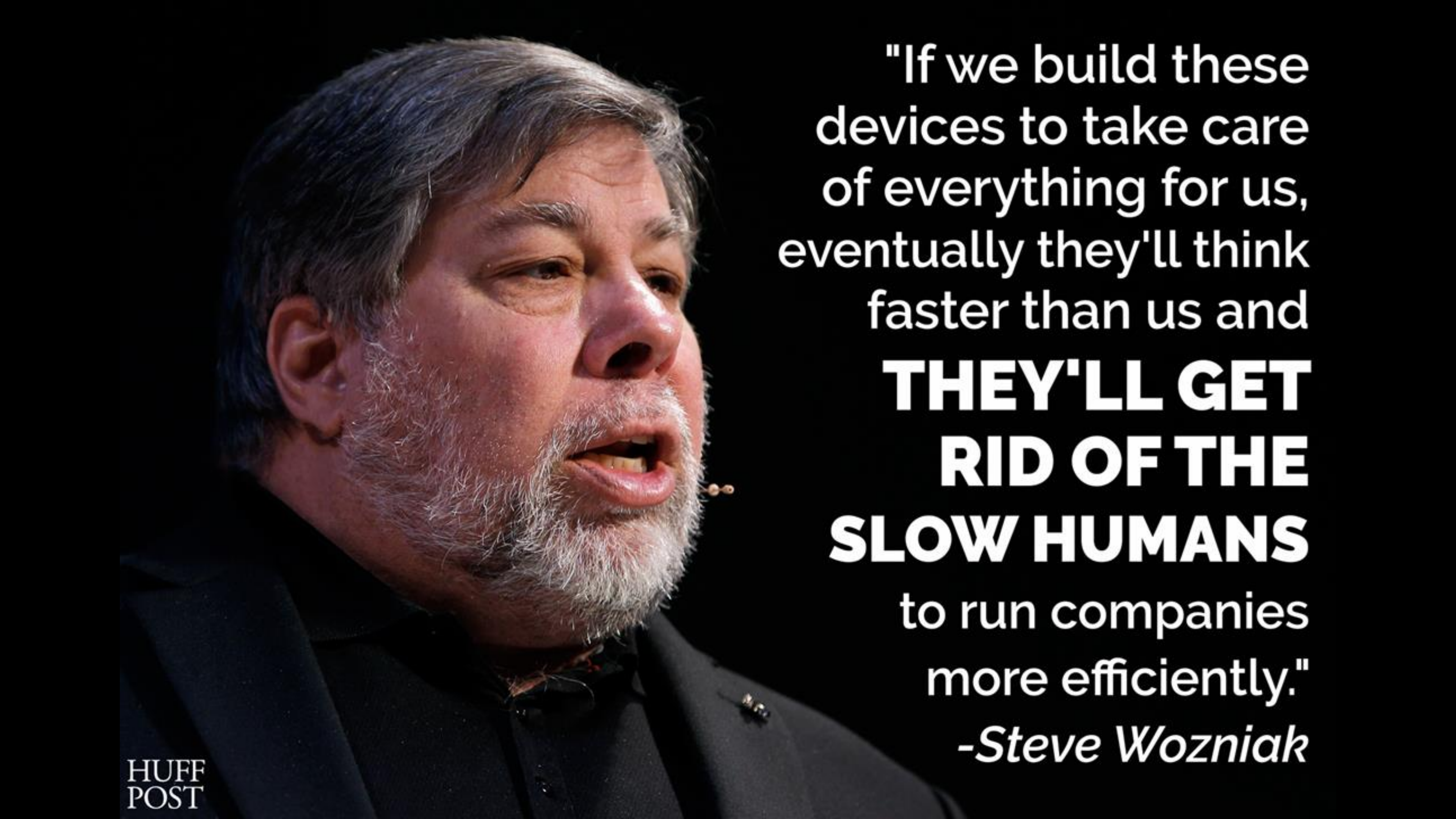
Now suppose X is actually A, then A must answer. It is A's

28 433

We are in the midst of an AI revolution

- Machine intelligence is progressing at an astounding rate
- Powered by Deep Learning algorithms that use huge amount of data to train complicated programs "Neural Networks"
- Can recognise faces and transcribe spoken sentences
- Can spot financial frauds
- Beat human in Go/Chess
- Enable self-driving cars
- Provide cancer diagnosis

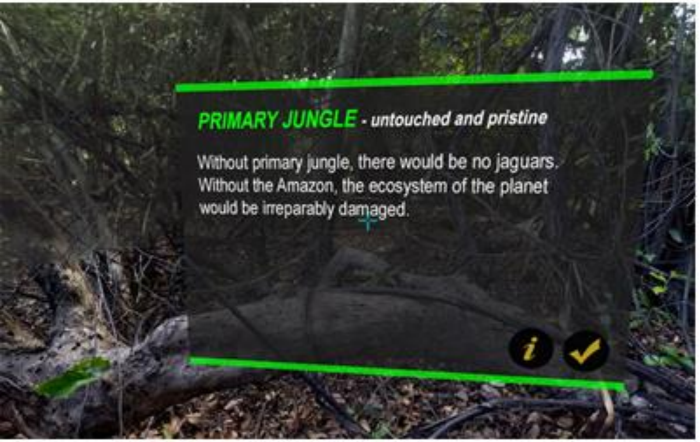


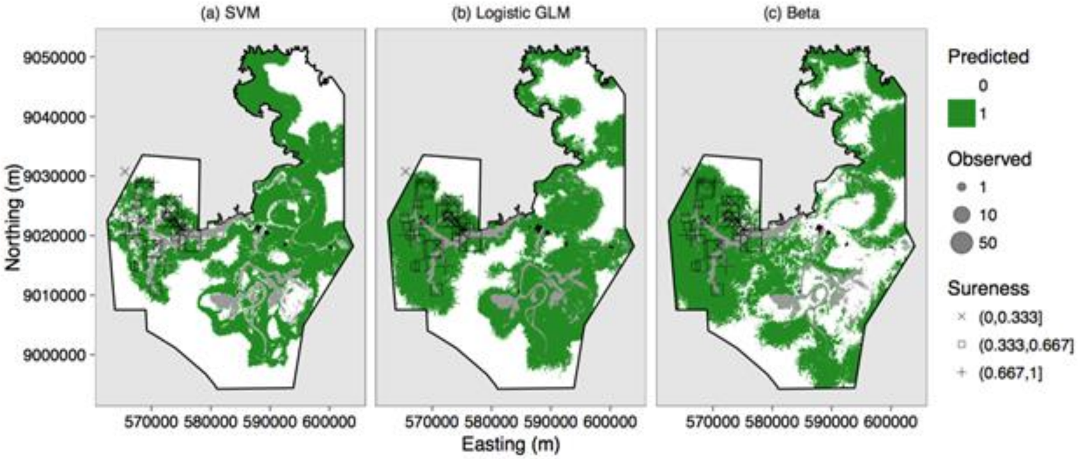

"If we build these devices to take care of everything for us, eventually they'll think faster than us and THEY'LL GET RID OF THE SLOW HUMANS to run companies more efficiently."
-Steve Wozniak
HUFF POST

Human Intelligence vs Artificial Intelligence

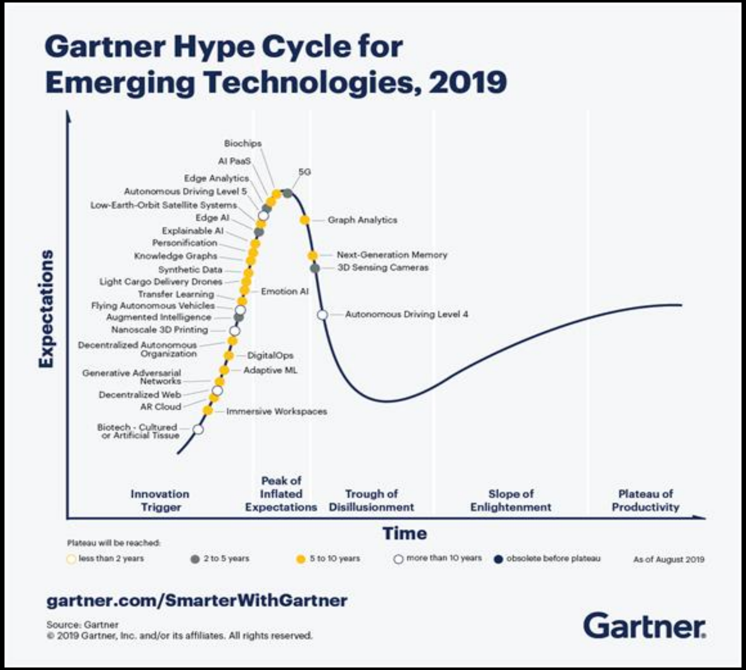
- Human Intelligence
  - Intuition, common sense, judgment, creativity, beliefs etc.
  - The ability to demonstrate their intelligence by communicating effectively.
  - Plausible reasoning and critical thinking.
  - Humans are fallible.
  - They have limited knowledge bases.
  - Information processing of serial nature proceed very slowly in the brain as compared to computers.
  - Humans are unable to retain large amounts of data in memory.
- Artificial Intelligence
  - Ability to simulate human behaviour and cognitive processes.
  - Capture and preserve human expertise.
  - Fast response. The ability to comprehend large about of data quickly.
  - No “common sense”.
  - Can not readily deal with “mixed” knowledge.
  - May have high development cost.
  - Raise legal and ethical concerns.

Saving Jaguars (VR, Gaming, GPUs and AI)





Emerging Technologies: the Hyper Cycle



Trough of Disillusionment

- 3D Sensing Cameras

Peak of Inflated Expectations

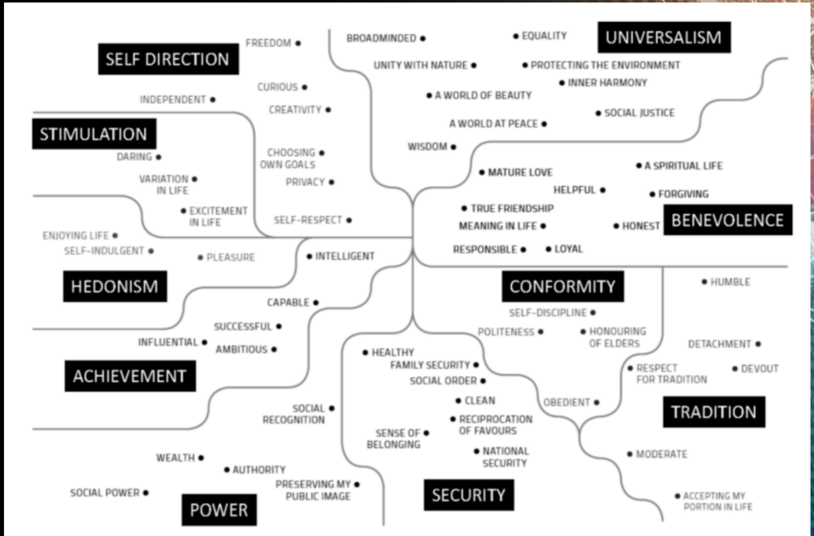
- 5G
- AI PaaS
- Edge Analytics

Innovation Trigger

- Emotion AI
- Adaptive ML
- Light Cargo Delivery Drones
- BioTech

Human Values in Software Engineering

- More diverse values and ethics need to be operationalised via Software Engineering
  - Not just security, safety & privacy
  - Need inclusion, diversity, responsibility, transparency, well-being, fairness, respect…
- XAI helps understanding and operationalising values in AI systems and improves trust



Mougouei *et. al* (2019), Operationalizing Human Values in Software: a Research Roadmap, FSE
Whittle *et. al* (2019), A Case for Human Values in Software Engineering, IEEE Software

Issues Impacting Trust in AI Systems

Bias in AI

- Amazon AI Recruiting tool
- Facial Recognition and skin colour
- Facebook Targeted Adverts
- US Healthcare Allocation
- Giggle

Visibility of Choice

Building Trust

- Why is AI biased?
- How do we evaluate bias?
- How do we illuminate bias?
- Can we avoid bias?

Data Integrity?

SIGGRAPH ASIA 2020 VIRTUAL

An Introduction to eXplainable AI

The Good, the Bad, and the Ugly

The Good: Artificial Intelligence (AI) technology can do amazing things in a wide range of domains (medicine, transport, robotics, games, etc.)

The Bad: These AIs are more and more built on technologies such as Machine Learning that are opaque, “black boxes”, and we generally can’t understand how or why the AI made the classification/decision/action that it did.

The Ugly: Stakeholders reject or distrust the AI (under-use), or blindly believe in the output of the AI without considering the consequences (over-reliance) (Pynadath 2018)

What is an Explanation?

- An answer to a question?
- A **story**?
- A conversation?
- A lecture?
- A report?
- A graph / chart / **visualization**?

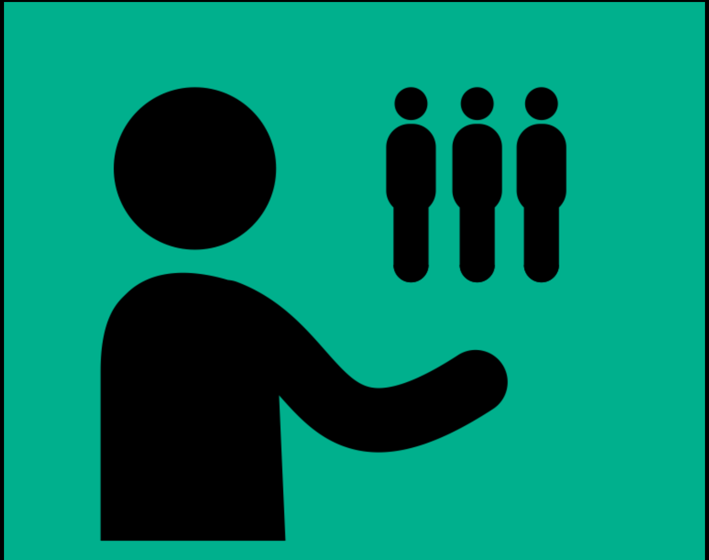
Explaining a game to new players (Photo credit: Paris Buttfield-Addison)

What is an Explanation?

Properties of an Explanation (Miller 2019)

- contrastive
  - i.e. in response to some counterfactual information (e.g. in response to "why did X happen instead of Y?")
- selected
  - i.e. from a range of almost infinite causes, we select (in a biased way) the most useful
- refer to causes, not probabilities
- social
  - i.e. presented as part of a conversation or **interaction**, in the context of the beliefs of explainer and explained

What is an Explanation?



Created by corpus delicti
from Noun Project

Characteristics of an Explanation
(Sridharan and Meadows 2019)

- representation abstraction
  - i.e. detail level of knowledge representation
  - e.g. “things in the room” vs “cups and books” vs “cup handle and cup base”
- communication specificity
  - i.e. selecting what spectrum of information to convey
  - e.g. describe the objects in the room, vs describe objects, current status of robot, the goal, etc.
- communication verbosity
  - i.e. how comprehensive the information is
  - e.g. “there was a child on the road” vs “there were 5 objects in front of the car, 4 cars and a child”

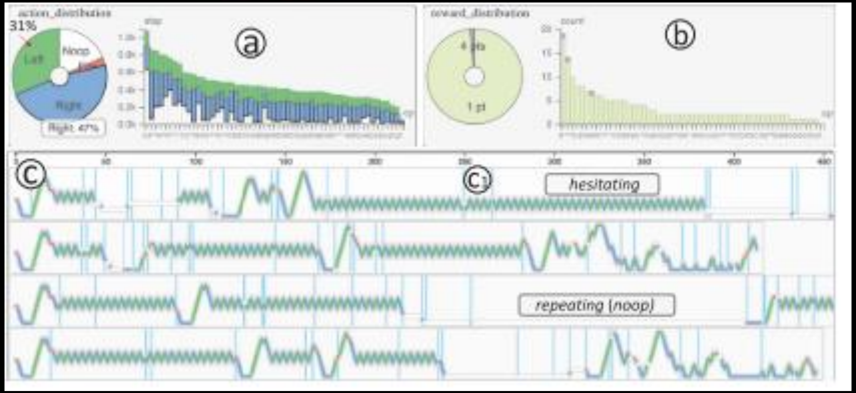
So What is XAI?

*"Given an audience, an explainable Artificial Intelligence is one that **produces details** or **reasons** to make its functioning clear or easy to **understand**."*
- (Arrieta et al. 2020)

- Why did you do that?
- Why not something else?
- When do you succeed?
- When do you fail?
- When can I trust you?
- How do I correct an error?

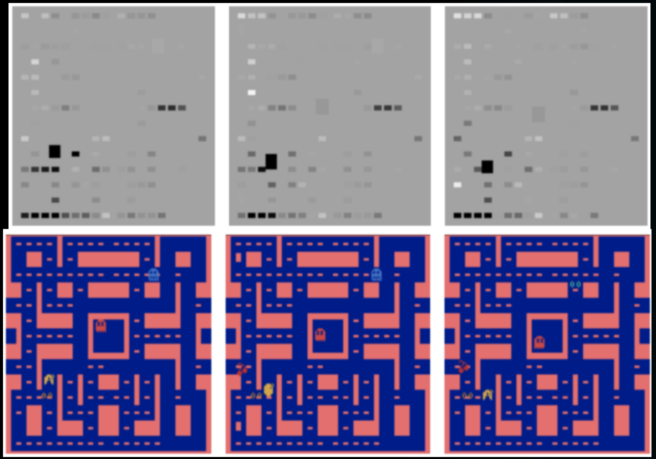
(Gunning 2017)

So What is XAI?



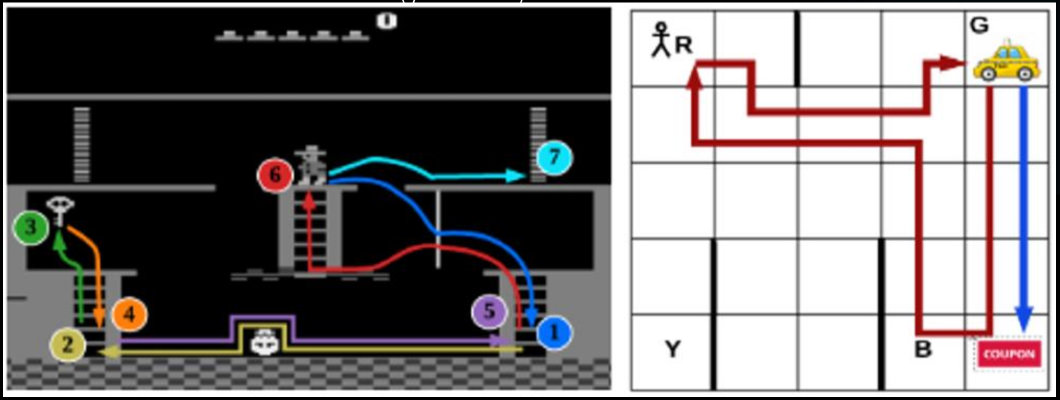
(Wang et al. 2019)

“Why not build barracks?”
“Because it is more desirable to do action build_supply_depot to have more supply depots as the goal is to have more destroyed units and destroyed buildings”.
(Madumal et al. 2019)


(Iyer et al. 2018)

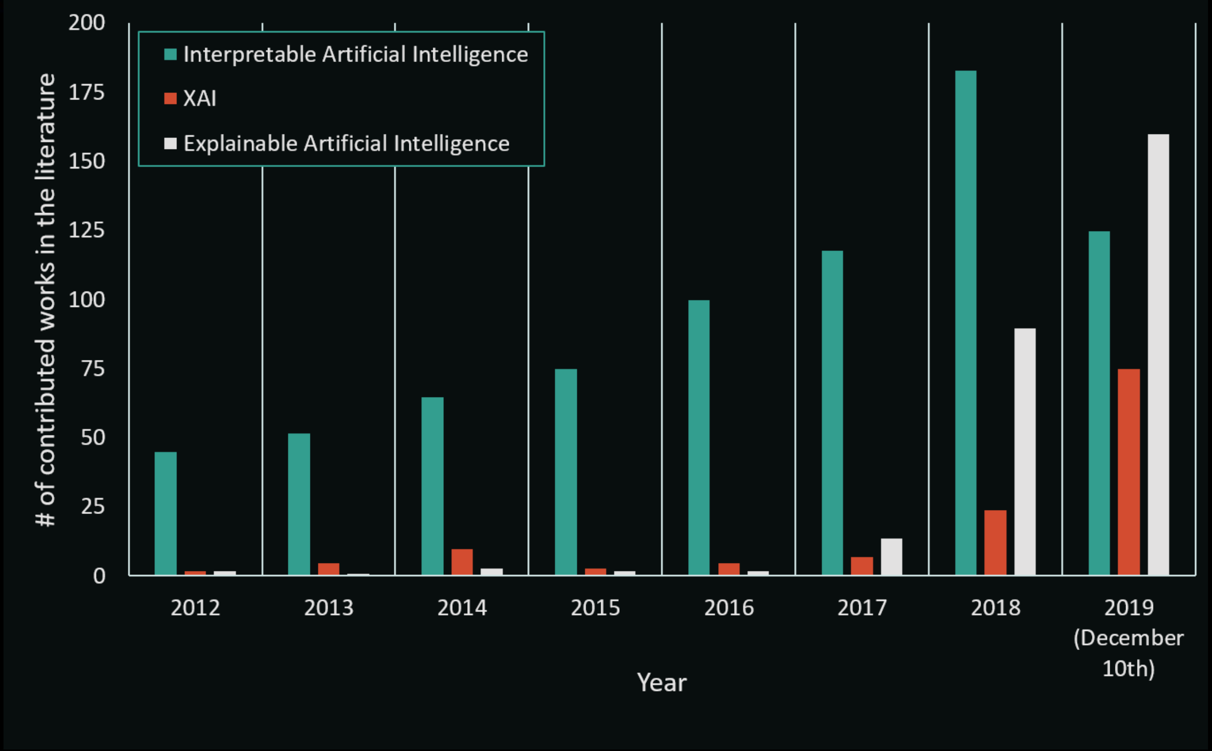
(Gunning 2017)

(Ehsan et al. 2019)



(Lyu et al. 2019)

The Focus has Shifted

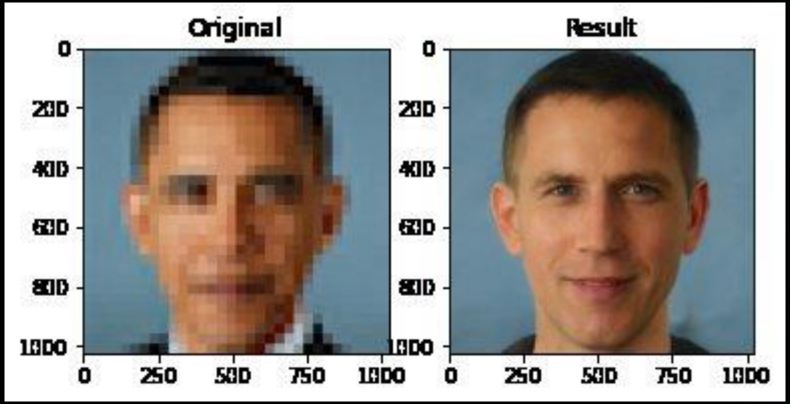


The Focus has Shifted

- **Google** "What-If" tool:
  - https://pair-code.github.io/what-if-tool/
  - Compare models
  - Visualize the importance of each feature value on each prediction
  - For any datapoint, find the most similar datapoint of a different classification
  - Compatible with TensorBoard, Jupyter, and Colab notebooks
- **Microsoft** Interpretability Framework:
  - https://docs.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-interpretability
  - *"empowers executives to understand "when deployed" how the model is working and how its decisions are treating and impacting people in real life."*
- **IBM** AI Fairness 360:
  - https://developer.ibm.com/open/projects/ai-fairness-360/
  - *"enables developers to use state-of-the-art algorithms to regularly check for unwanted biases from entering their machine learning pipeline and to mitigate any biases that are discovered."*

Why XAI?

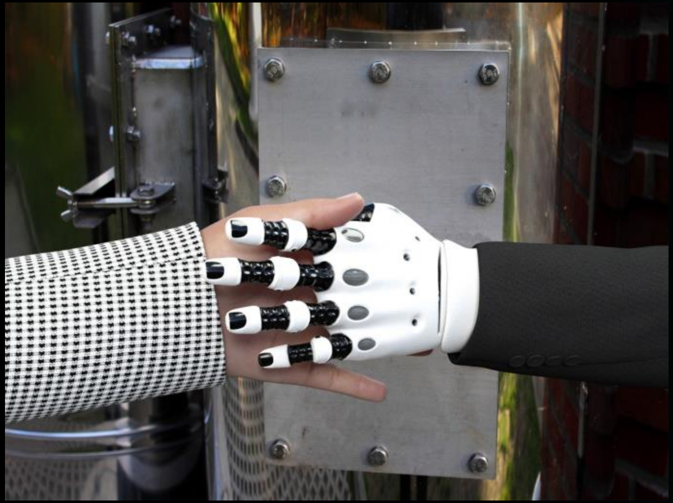
- Stepping stone in the pursuit of "Artificial General Intelligence" (Goertzel 2007)
  - A truly intelligent machine should be able to rationalize and explain its actions and decisions
- Debugging and Elimination of Bias
  - If our AI is making strange decisions, XAI visualizations may make mistakes obvious (e.g. incorrectly labelled training data, lack of diverse examples, unintended focus on specific features, etc.)
- Human-in-the-loop



https://twitter.com/Chicken3gg/status/1274314622447820801

Why XAI?

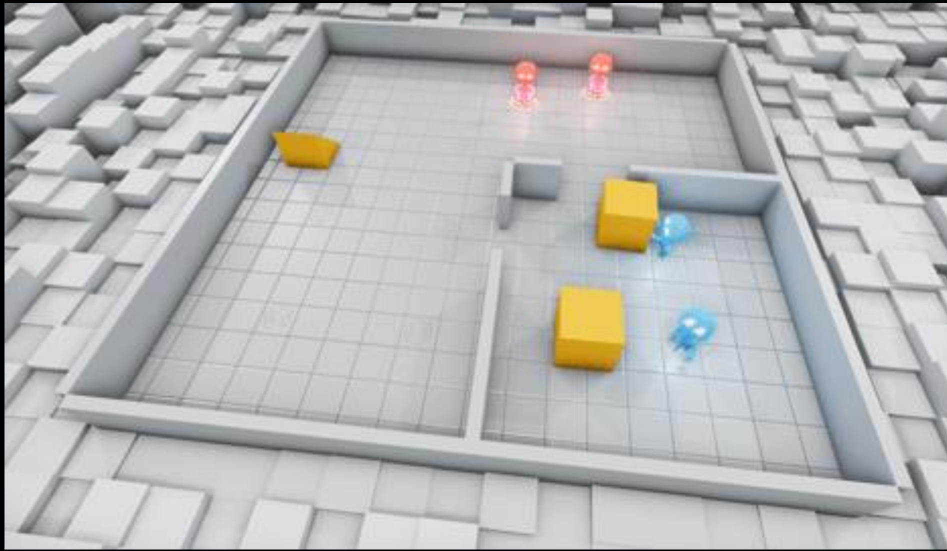
- Trust and Safety
  - Autonomous vehicles and robotics (Araiza-Illan 2019)
    - Why a robot took the action it did
    - What factors it considered
    - Trusting that the robot is indeed making intelligent and safe decisions
  - Building a rapport with the robot
    - Can make working with it more efficient
  - Diagnosing problems causing an incident
    - Query why it took certain actions
    - Prevent further incidents
    - Accountability
    - Insurance or legal claims
      - Car at fault? Fault in the decision making?
      - Third party at fault?



Why XAI?

- Decision Making and Discovery of New Policies
  - Identify new strategies or policies
    - E.g. hide-and-seek where agents learned strategies to exploit the physics system of the game to overcome walls that were not supposed to be passed (Baker et al. 2019)
  - Extracting how these strategies were learned, or under what circumstances these strategies were learned could result in useful new knowledge for decision making or optimisation.

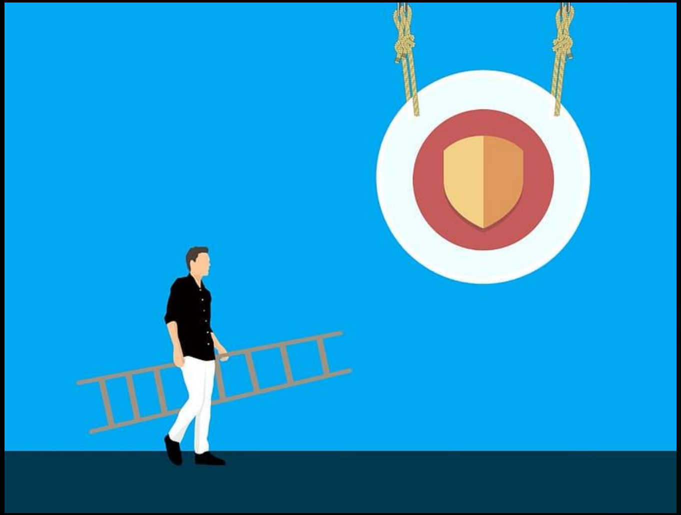
"[...] people and computers can both play chess, it is far from clear whether they do it the same way" - Zhuang (2017)



Multi-Agent Hide and Seek (from Baker et al. 2019)
https://www.youtube.com/watch?v=kopoLzvh5jY

Challenges in the XAI Space

- Creating model-agnostic approaches
- Defining an explanation
- Generating human-readable explanations
  - Plus interfaces/language for querying an AI
- Tailoring output based upon level of human knowledge (domain experts vs lay people)
- Scalability of solutions
- Measuring the effectiveness of explanations?
- Explanations of temporal AI systems (i.e. Deep Reinforcement Learning)
  - E.g. explanations over a set of actions in sequence, or actions that only have an interesting effect after a delay



Stages of AI Explainability

Pre-modelling Explainability

- Understand/describe data used to develop models

Explainable Modelling

- Develop inherently more explainable models

Post-modelling Explainability

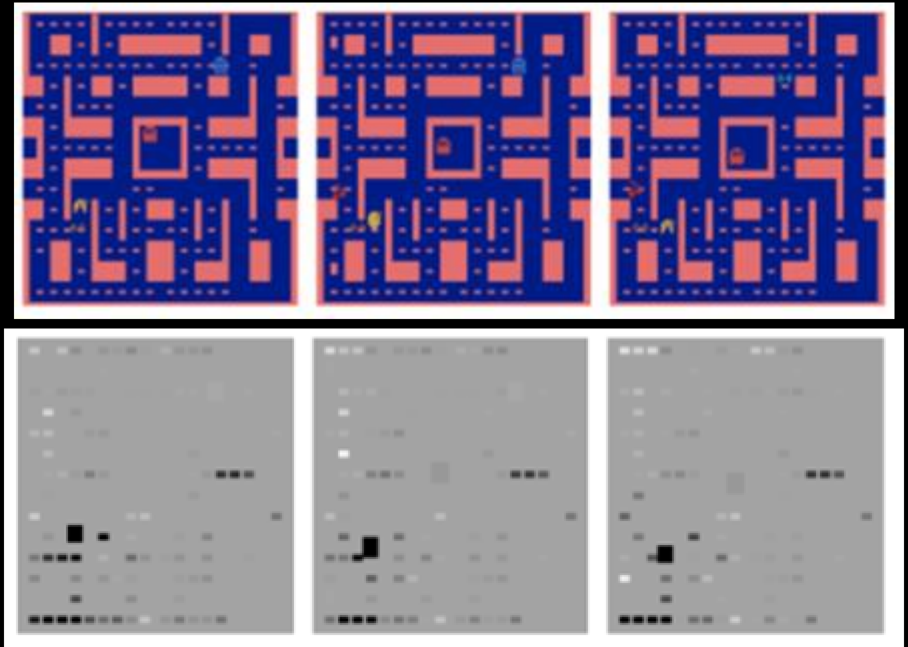
- Extract **explanations** to describe pre-developed models

Adapted from Khaleghi (2019), used with permission

Saliency Maps - An Example of Visualization for XAI

(Iyer et al. 2018, p. 148)

- Highlight areas of the input image that were of importance to the outcome

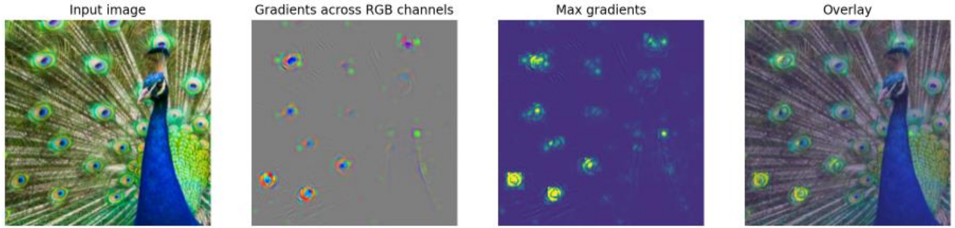


Saliency Maps

Exercise 1

What makes peacock a peacock...?

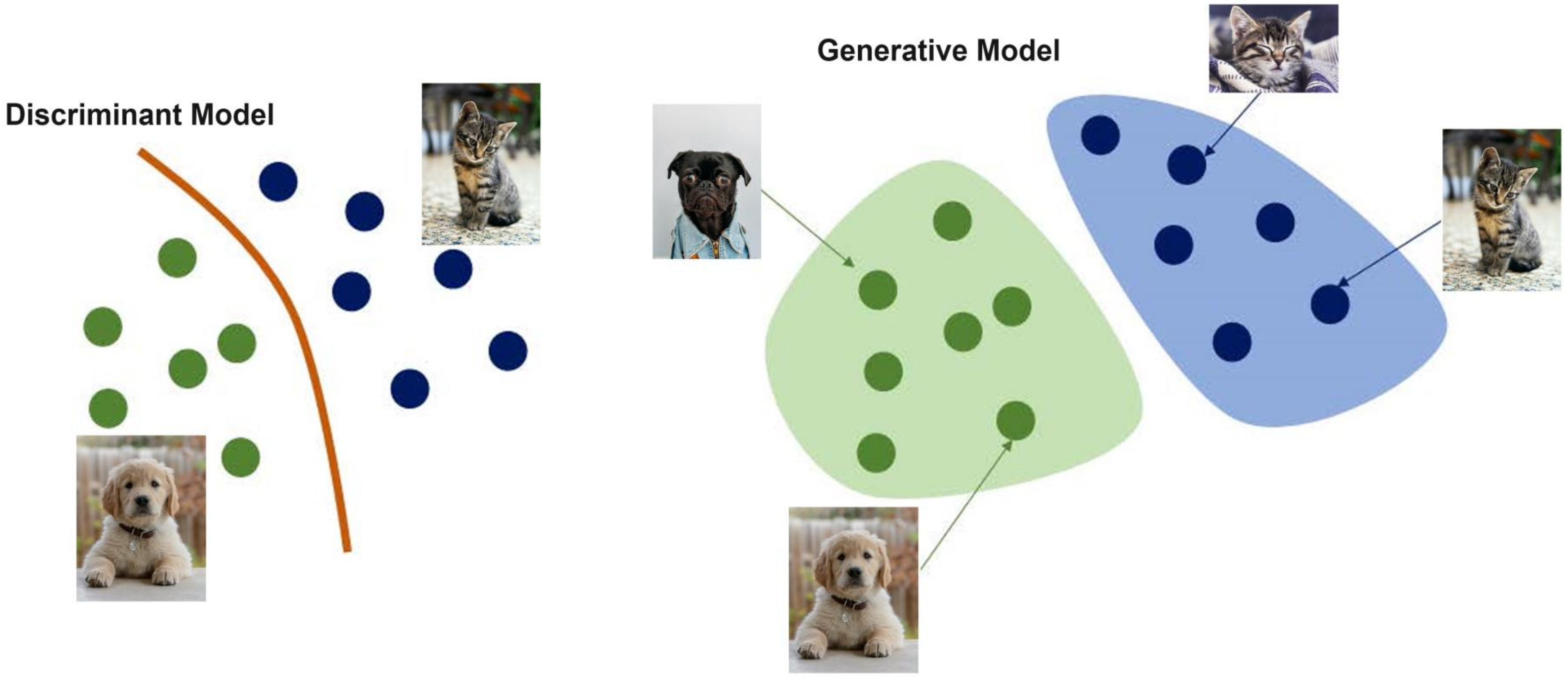
```
[ ] peacock = apply_transforms(load_image('/content/images/peacock.jpg'))
    backprop.visualize(peacock, 84, guided=True, use_gpu=True)
```



SIGGRAPH ASIA 2020 VIRTUAL

XAI by Default - Interactive AI and Generative Models

Interacting with AI - Generative Models



XAI & Generative Networks

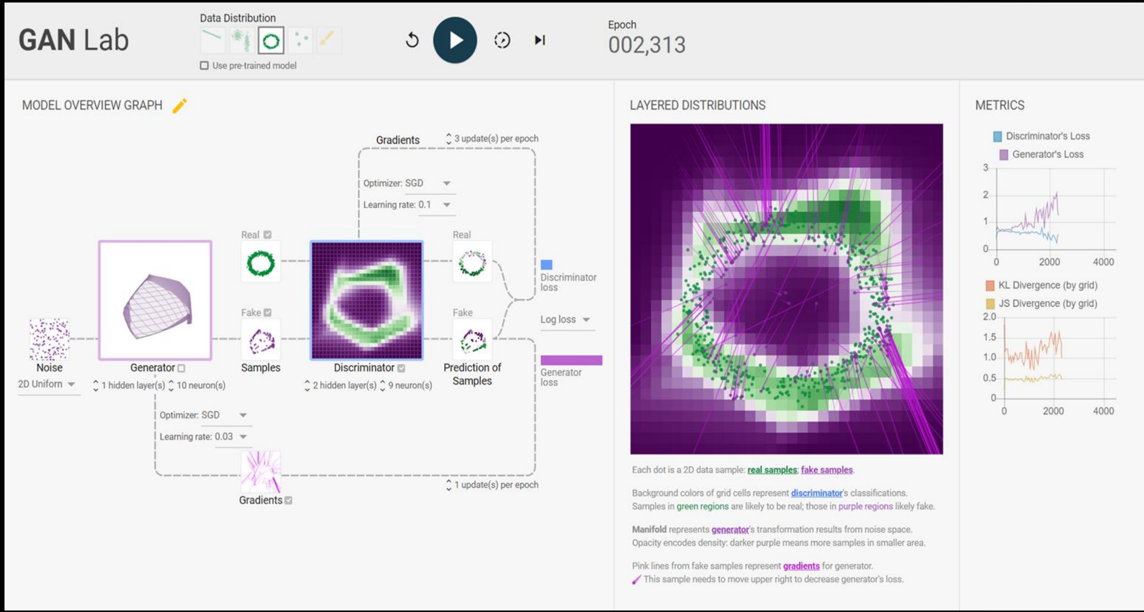
- As model is being trained.
- Provides understanding how the model evolves.
- Interactive control over model parameters.
- Visual Analytics tools to provide interactive feedback.

Photo by Josue Isai Ramos Figueroa on Unsplash

XAI & Generative Networks

Example - Gan Lab

- Shows how model evolves over time and how changing parameters impacts results.
- Simple User-Interface.
- Some prior knowledge needed.
- Great pedagogic tool.



Adapted from Kahng (2018), used with permission

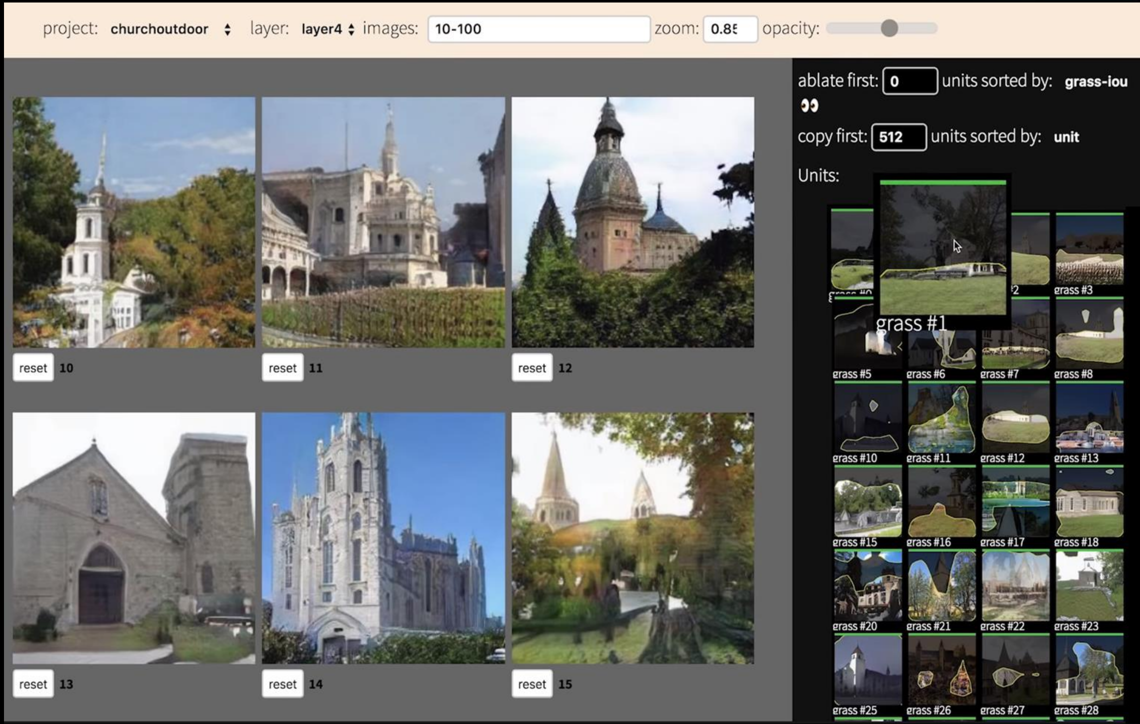
XAI & Generative Networks

- After model has been trained.
- XAI tools can provide control over the network in a human-understandable way.
- Attempts to segment into user understandable "objects".
- Interactive control over model parameters.
- Visual Analytics tools to provide interactive feedback.

XAI & Generative Networks

Example - GAN Dissection

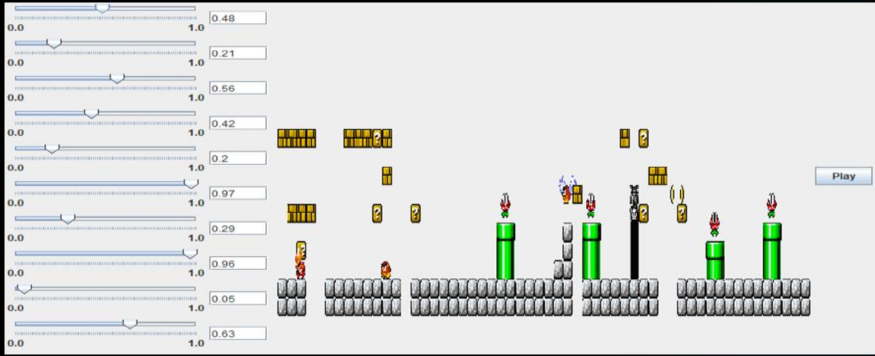
- Allows user to mark up images, creating a mapping from object concepts to corresponding neural units.
- Simple User-Interface.
- No prior knowledge needed.
- Immediately useful for further model refinement, end-users and general model understanding



Bau (2018), used with permission

Unsupervised Feature Extraction

- There are many approaches to automatic representation disentanglement (e.g. InfoGAN).
- Can return meaningful, controllable features to users.



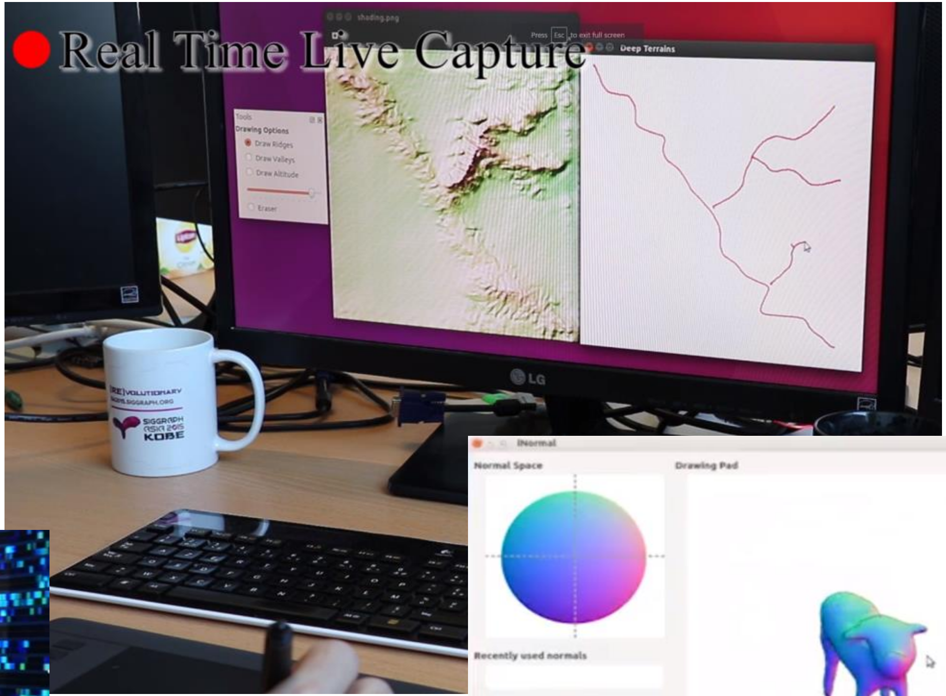
Schrum (2020), used with permission

(a) Rotation (b) Width

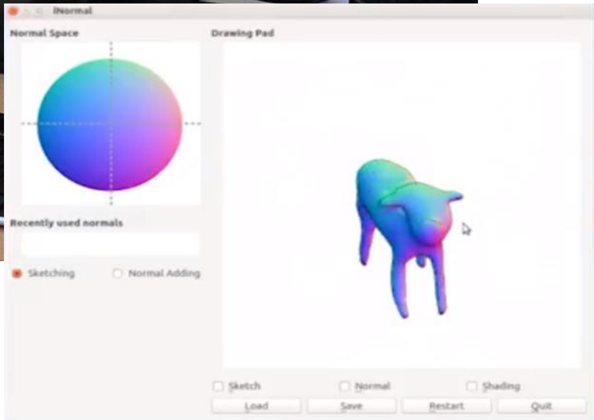
Chen (2016), used with permission

Interaction Design for Better Human-AI Collaboration

- When building User-Interfaces for AI systems, it is critical to keep the end-user in mind and preferably part of the UI design team.
- Well-designed, intuitive UIs that expose the features that the end-user expects leads to increased trust.
- It also allows the end-user to explore the systems capabilities and quickly understand what it can and cannot achieve.



Guérin (2017)



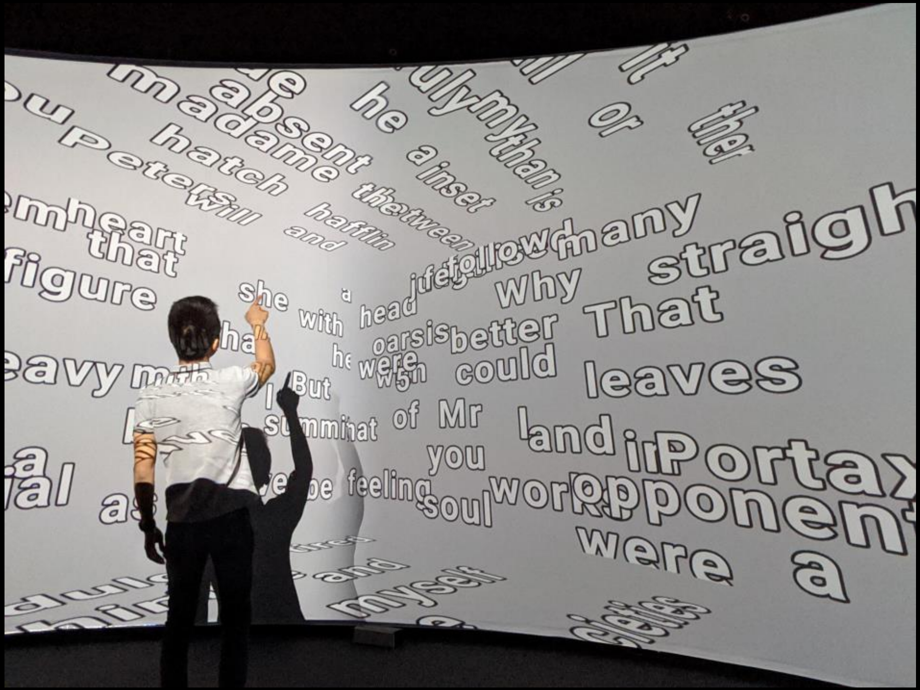
Su (2018)

Interactive Analytics Example

Exercise 2

Narrative solutions for XAI

"Artificial Intelligence is almost a humanities discipline. It's really an attempt to understand human intelligence and human cognition."
-Sebastian Thrun



What is storytelling? What is narrative?

- Storytelling is to tell a story.
- Story is a fuzzy term: it is an event, a moment, etc.
- A narrative is one way of telling or examining a story.
- How are structure, characters, techniques and events ordered and utilised?
- Another way to think of it is the terms “fabula” and “szujet”, which refer to the raw elements of the story and the order they are told in, respectively (Prince, 1995).

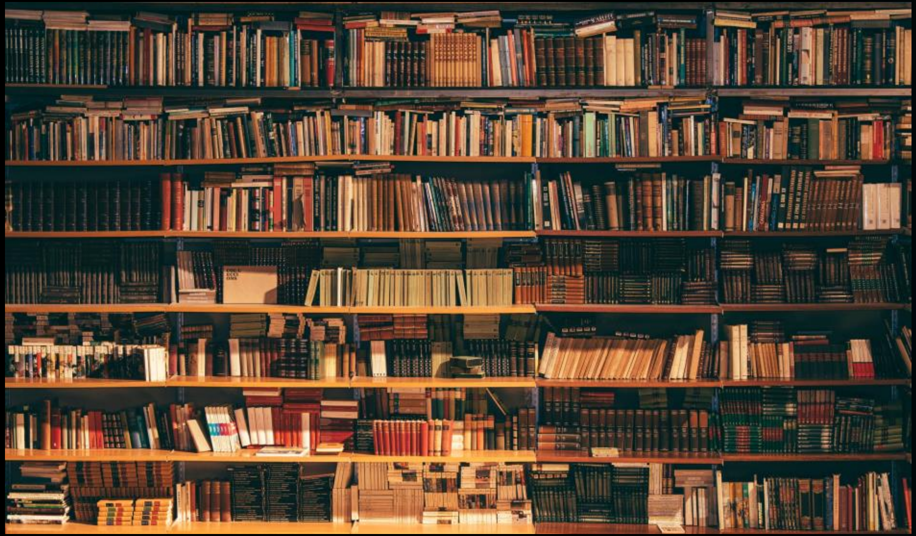


Image courtesy of Alfons Morales (2017) via unsplash.com

Story vs Narrative

Story:
Alan Turing’s life and legacy

- A vast series of moments
- Messy and linear
- Many “boring” bits.
- Cannot be told easily
- Involves people.

TURING TEST PICTURES | DIRECTED BY JOHN FILMMAKER | WRITTEN BY KATHRYN WORDSMITH | STARRING LISA THESPIAN & RODNEY MOTIVATION | MOSTLY BASED ON A TRUE STORY | PRODUCED BY MONEYBAGS BROS

Narrative:
Alan Turing: The movie

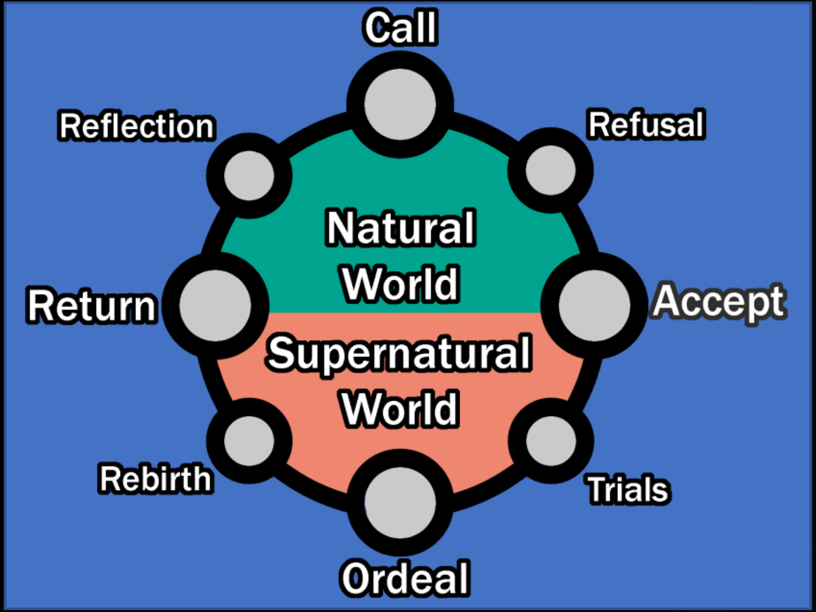
- Focuses on events
- Can skip through time
- Embellished and dramatic
- An imperfect solution to a difficult problem.
- Involves characters.

Can we use narratives to tell an AI’s story?

Narrative structures

The Monomyth/Hero's Journey

- Joseph Campbell's widely referenced structure (1949)
- Suggests all narratives follow one structure
- Criticized for being far too broad/not doing all it says it does



Narrative structures

Folktale Morphology

- Created by Vladamir Propp (1928/1968)
- Formulaically explains folktale structure
- Focused on characters being able to perform certain actions (e.g., a villain could attempt to spy on the kingdom, hurt a family member, etc.)

$$y^1\beta^1\delta^1A^1C\uparrow\begin{pmatrix}[DE^1\,neg.F\,neg.]\\ d^7E^7F^9\end{pmatrix}G^4K^1\downarrow[Pr^1D^1E^1F^9=Rs^4]^8$$

"The Swan-Geese"

Propp, 1968, pp. 96-99)

Narrative structures

Rhetorical Structure Theory

- Conceived for computational linguistics (Taboada & Mann, 2006)
- Used more and more for narratives specifically (Nakasone & Ishizuka 2006; Gaeta et al., 2014).
- Views bodies of text as ‘units” with relations to one another (background, elaboration, causes, etcetera)

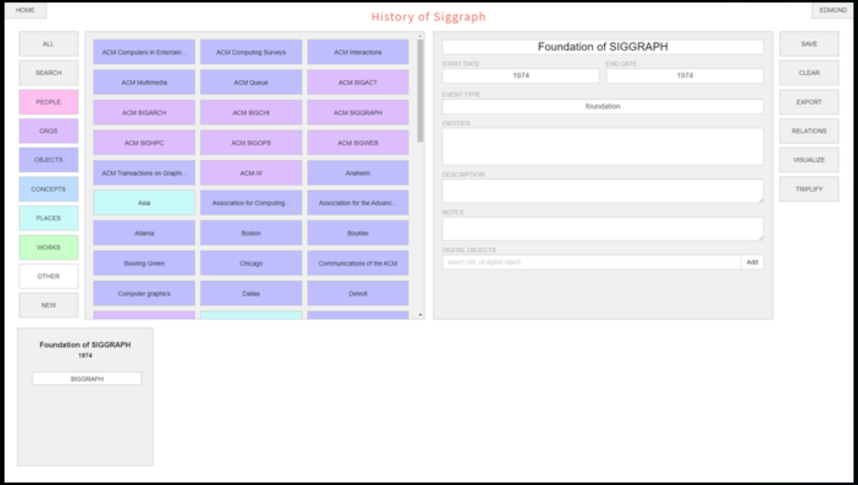
Narrative Visualisation Analytics Toolbox (Image courtesy of EPICentre and Defense Science and Technology)

Narrative structures

...And many more!

Narrative elements

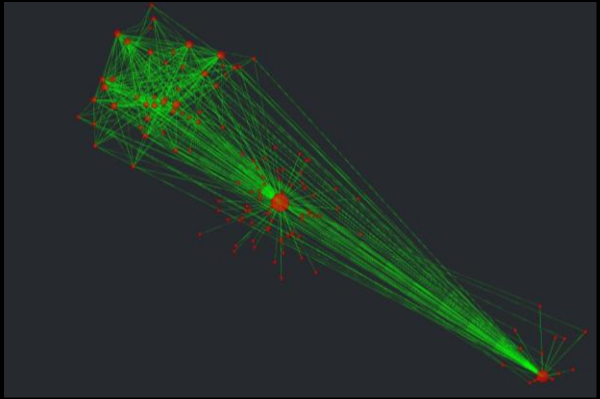
- **A beginning, a middle, an end…**
  - Events must be in some way temporally arranged to constitute a narrative.
- **Characters**
  - The entities that drive the narrative. Referred to as “spheres of action” by Chatman (1978) due to being defined by their agency.
- **Techniques/Modes**
  - The various smaller “parts” of a narrative.
  - Includes decisions such as what tense to use and whether the narrator is speaking in first or third person
  - Also includes techniques such as the use of metaphors, the use of flashbacks, dialogue, and more.



NBVT (Meghini et al, 2020) (Image courtesy of ISTI-CNR)

XAI Narratives for Human-In-The-Loop

- "We hypothesize that narrative explanation will be more easily understood by non-expert human operators of artificial intelligence since the human mind is tuned for narrative understanding." (Riedl, 2016, p. 2)



Why use narratives?

Explanation	Narrative
Contrastive	Exposition, action
Selected	Flashforwards, flashbacks, ellipsis, amplification
Refer to causes, not probabilities	Action, characters
Social	Draws on cultural norms, symbolisms
Representation abstraction	Metaphors, Personification, POV
Communication specificity	Metaphors, Personification, POV
Communication verbosity	Exposition, Tense, Pathos

Using Narratives

Think about the last time you had to explain something...

- Were there characters?
- A point of view?
- Were details omitted or embellished?
- Did you use metaphors or allegories to describe certain elements?
- Did you skip through time?
- Humans are natural storytellers. We've been reading – and telling – stories for most of our lives.

Image courtesy of Steffen Wienberg (2020), via unsplash.com

A Narrative is just a particular type of Explanation...

...but Uncertainty is More Important to AI than Explainability

Responding to User Queries

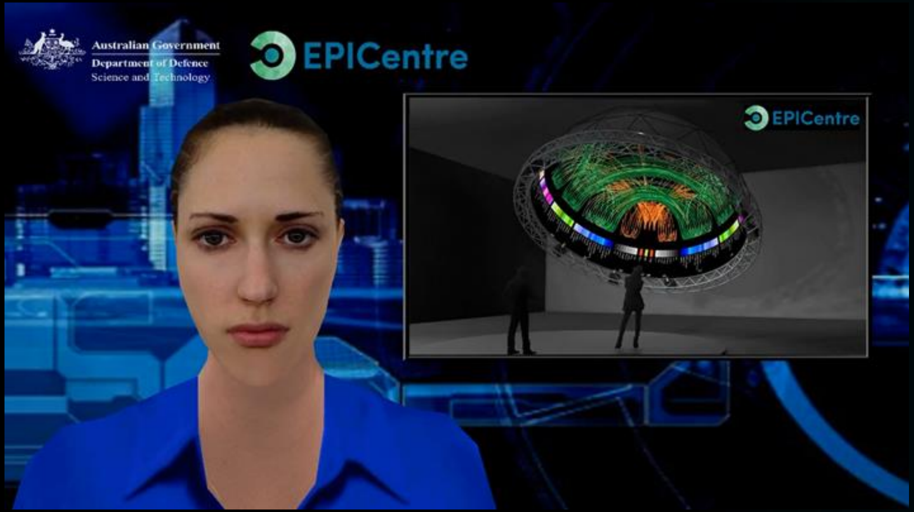
Think about the last time you had to explain something...

- Were there characters?
- A point of view?
- Were details omitted or embellished?
- Did you use metaphors or allegories to describe certain elements?
- Did you skip through time?
- Humans are natural storytellers. We've been reading – and telling – stories for most of our lives.

“Question Mark” by ranmilani is licensed under CC BY 2.0

Lessons from NarVis

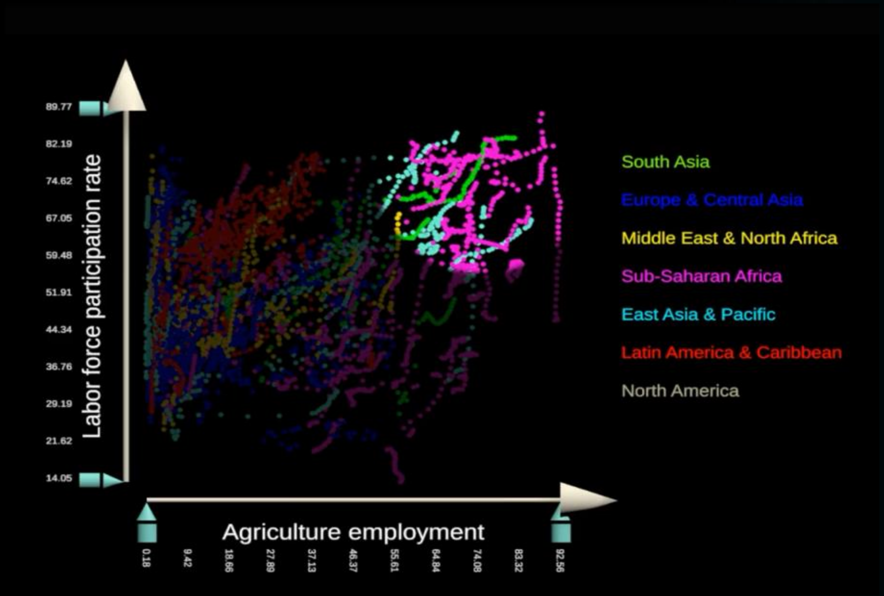
- Narrative Visualisation (NarVis) is a burgeoning area of data communication.
- Like XAI, NarVis is data-driven, seeking to better communicate data.
- Could NarVis and XAI Narratives develop in tandem?



Narrative Visualisation Analytics Toolbox (Image courtesy of EPICentre and Defense Science and Technology)

Benefits of Narrative

- Can help build user's empathy
- Narrative vagueness oddly builds trust
- Bootstraps on pre-established models and concepts, so is useful for non-experts
- (Krause & Rucker, 2019)



Narrative Visualisation Analytics Toolbox (Image courtesy of EPICentre and Defense Science and Technology)

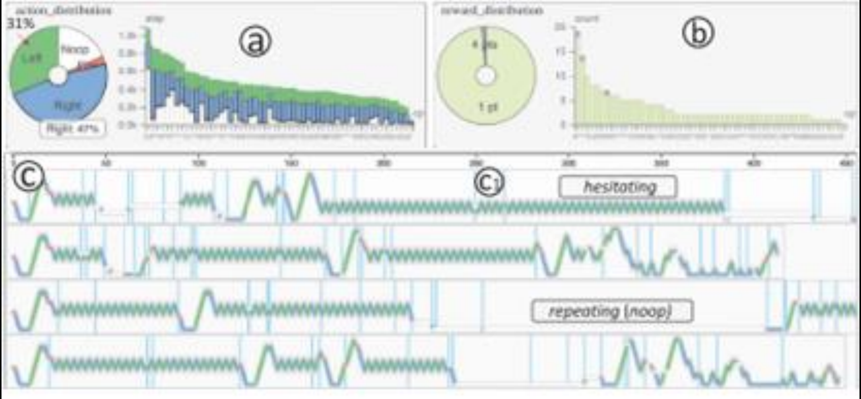
SIGGRAPH ASIA 2020 VIRTUAL

Dangers of Narrative

- Potential for misrepresentation
- Persuasiveness can be misused
- Imagery, concepts and metaphors may bring their own baggage
- (Krause & Rucker, 2019)

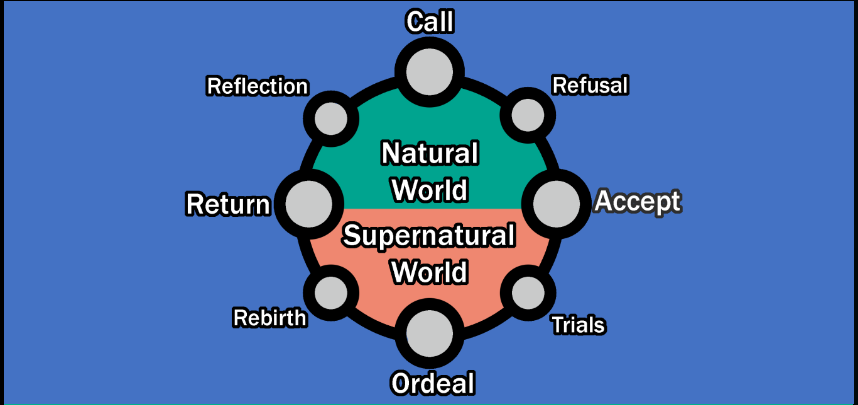
Narrative Visualisation Analytics Toolbox (Image courtesy of EPICentre and Defense Science and Technology)

Narrative workflow



Data collection

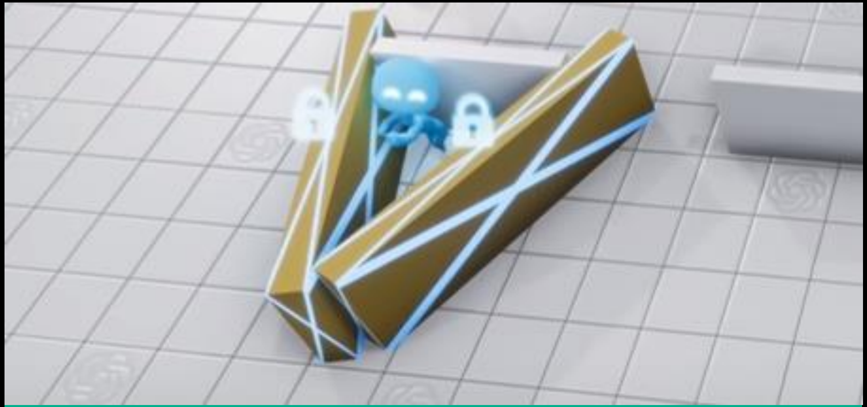
- XAI processes
- Collecting the AI's story
- Should they be logging their data in a certain way?



Data narrativisation

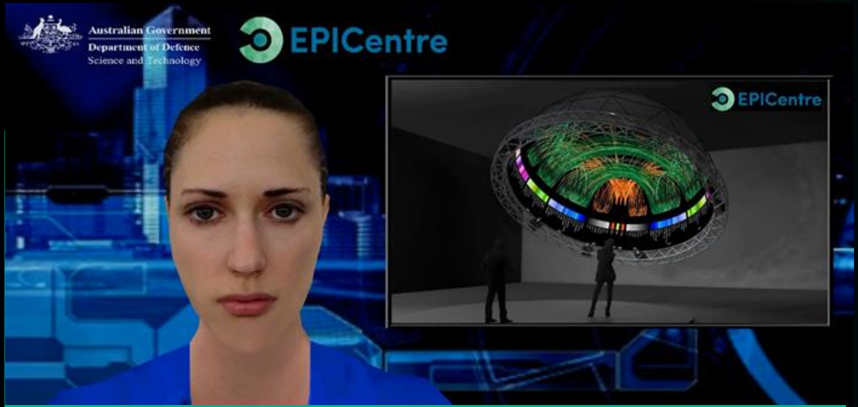
- How to order events
- Causality
- Who are the "actors" in this story? Or the key events?

Narrative workflow



Data abstraction/visualisation

- Assigning metaphor/symbolism
- Representation and curation
- What parts of the AI's story are better told through visualisation?



Narration/communication

- Modification for audience
- Temporal structure
- Has anything been lost in translation?

Wrap-Up

Challenges ahead

XAI

- When to use narratives
- What questions to answer

Narrative and Visualisation

- Avoiding bias
- Non-expert friendly
- Choosing techniques
- Harmonising with visualisation

NLP

- Quality and complexity

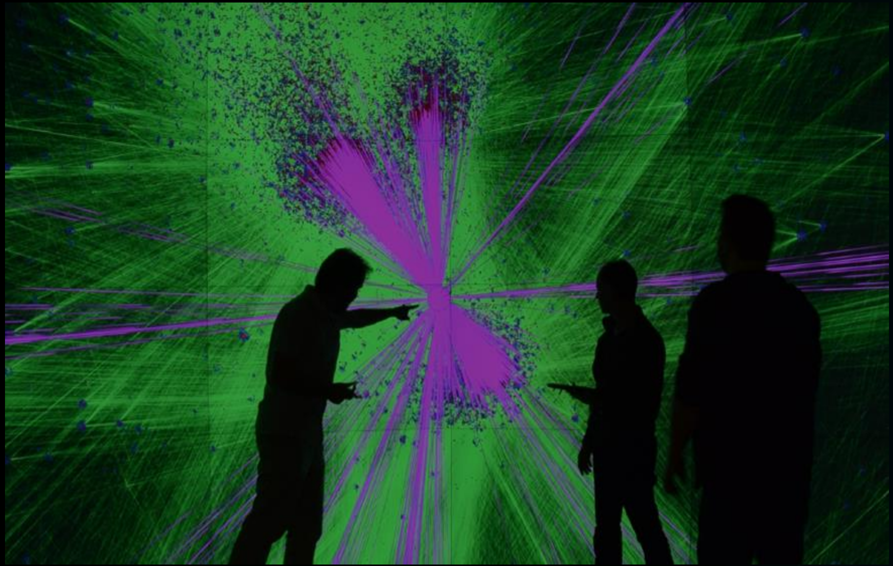


Image courtesy of EPICentre

References

Araiza-Illan D., Eder K. (2019) Safe and Trustworthy Human-Robot Interaction. In: Goswami A., Vadakkepat P. (eds) Humanoid Robotics: A Reference. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-6046-2_131

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Chatila, R. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82-115.

Baker, B., Kanitscheider, I., Markov, T., Wu, Y., Powell, G., McGrew, B., & Mordatch, I. (2019). Emergent tool use from multi-agent autocurricula. arXiv preprint arXiv:1909.07528.

Bau, D., Zhu, J. Y., Strobelt, H., Zhou, B., Tenenbaum, J. B., Freeman, W. T., & Torralba, A. (2018). Gan dissection: Visualizing and understanding generative adversarial networks. arXiv preprint arXiv:1811.10597.

Campbell, J. (2008). The hero with a thousand faces (Vol. 17). New World Library.

Chatman, S. B. (1980). Story and discourse: Narrative structure in fiction and film. Cornell University Press.

Chen, X., Duan, Y., Houthooft, R., Schulman, J., Sutskever, I., & Abbeel, P. (2016). Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In Advances in neural information processing systems (pp. 2172-2180).

Ehsan, U., Tambwekar, P., Chan, L., Harrison, B., & Riedl, M. O. (2019, March). Automated rationale generation: a technique for explainable AI and its effects on human perceptions. In Proceedings of the 24th International Conference on Intelligent User Interfaces (pp. 263-274).

Gaeta, A., Gaeta, M., & Guarino, G. (2014, September). RST-based methodology to enrich the design of digital storytelling. In 2014 International Conference on Intelligent Networking and Collaborative Systems (pp. 720-725). IEEE.

Goertzel, B. (2007). Artificial general intelligence (Vol. 2). C. Pennachin (Ed.). New York: Springer.

Guérin, É., Digne, J., Galin, É., Peytavie, A., Wolf, C., Benes, B., & Martinez, B. (2017). Interactive example-based terrain authoring with conditional generative adversarial networks. Acm Transactions on Graphics (TOG), 36(6), 1-13.

Gunning, D. (2017). Explainable artificial intelligence (xai). Defense Advanced Research Projects Agency (DARPA), nd Web, 2, 2.

Iyer, R., Li, Y., Li, H., Lewis, M., Sundar, R., & Sycara, K. (2018, December). Transparency and explanation in deep reinforcement learning neural networks. In Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society (pp. 144-150).

Kahng, M., Thorat, N., Chau, D. H. P., Viégas, F. B., & Wattenberg, M. (2018). Gan lab: Understanding complex deep generative models using interactive visual experimentation. IEEE transactions on visualization and computer graphics, 25(1), 1-11.

Krause R.J. & Rucker. D D. (2019). Strategic Storytelling: When Narratives Help Versus Hurt the Persuasive Power of Facts. Personality and Social Psychology Bulletin, 46(2), 216-227. doi: 10.1177/0146167219853845.

Lyu, D., Yang, F., Liu, B., & Gustafson, S. (2019, July). SDRL: interpretable and data-efficient deep reinforcement learning leveraging symbolic planning. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 33, pp. 2970-2977).

Madumal, P., Miller, T., Sonenberg, L., & Vetere, F. (2019). Explainable reinforcement learning through a causal lens. arXiv preprint arXiv:1905.10958.

Meghini, C., Bartalesi, V., Metilli, D., & Benedetti, F. (2020). Narrative Building and Visualising Tool (NBVT). https://dlnarratives.eu/tool.html

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. Artificial Intelligence, 267, 1-38.

Mougouei, D., Perera, H., Hussain, W., Shams, R., & Whittle, J. (2018, October). Operationalizing human values in software: a research roadmap. In Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering (pp. 780-784).

Nakasone, A., & Ishizuka, M. (2006, December). SRST: A storytelling model using rhetorical relations. In International Conference on Technologies for Interactive Digital Storytelling and Entertainment (pp. 127-138). Springer, Berlin, Heidelberg.

Prince, G. (2019). Narratology. In Oxford Research Encyclopedia of Literature.

Propp, V. (2010). Morphology of the Folktale (Vol. 9). University of Texas Press.

Pynadath, D. V., Barnes, M. J., Wang, N., & Chen, J. Y. (2018). Transparency communication for machine learning in human-automation interaction. In Human and Machine Learning (pp. 75-90). Springer, Cham.

Riedl, M. O. (2019). Human-centered artificial intelligence and machine learning. Human Behavior and Emerging Technologies, 1(1), 33-36.

Schrum, J., Gutierrez, J., Volz, V., Liu, J., Lucas, S., & Risi, S. (2020). Interactive evolution and exploration within latent level-design space of generative adversarial networks. In Proceedings of the 2020 Genetic and Evolutionary Computation Conference (GECCO '20). Association for Computing Machinery, New York, NY, USA, 148–156. DOI:https://doi.org/10.1145/3377930.3389821

Sridharan, M., & Meadows, B. (2019). Towards a Theory of Explanations for Human–Robot Collaboration. KI-Künstliche Intelligenz, 33(4), 331-342.

Su, W., Du, D., Yang, X., Zhou, S., & Fu, H. (2018). Interactive sketch-based normal map generation with deep neural networks. Proceedings of the ACM on Computer Graphics and Interactive Techniques, 1(1), 1-17.

Taboada, M., & Mann, W. C. (2006). Rhetorical structure theory: Looking back and moving ahead. Discourse studies, 8(3), 423-459.

Wang, J., Gou, L., Shen, H. W., & Yang, H. (2018). Dqnviz: A visual analytics approach to understand deep q-networks. IEEE transactions on visualization and computer graphics, 25(1), 288-298.

Whittle, J., Ferrario, M. A., Simm, W., & Hussain, W. (2019). A Case for Human Values in Software Engineering. IEEE Software.

Zhuang, Y. T., Wu, F., Chen, C., & Pan, Y. H. (2017). Challenges and opportunities: from big data to knowledge in AI 2.0. Frontiers of Information Technology & Electronic Engineering, 18(1), 3-14.