Sztuczna Inteligencja

Uczenie maszynowe III

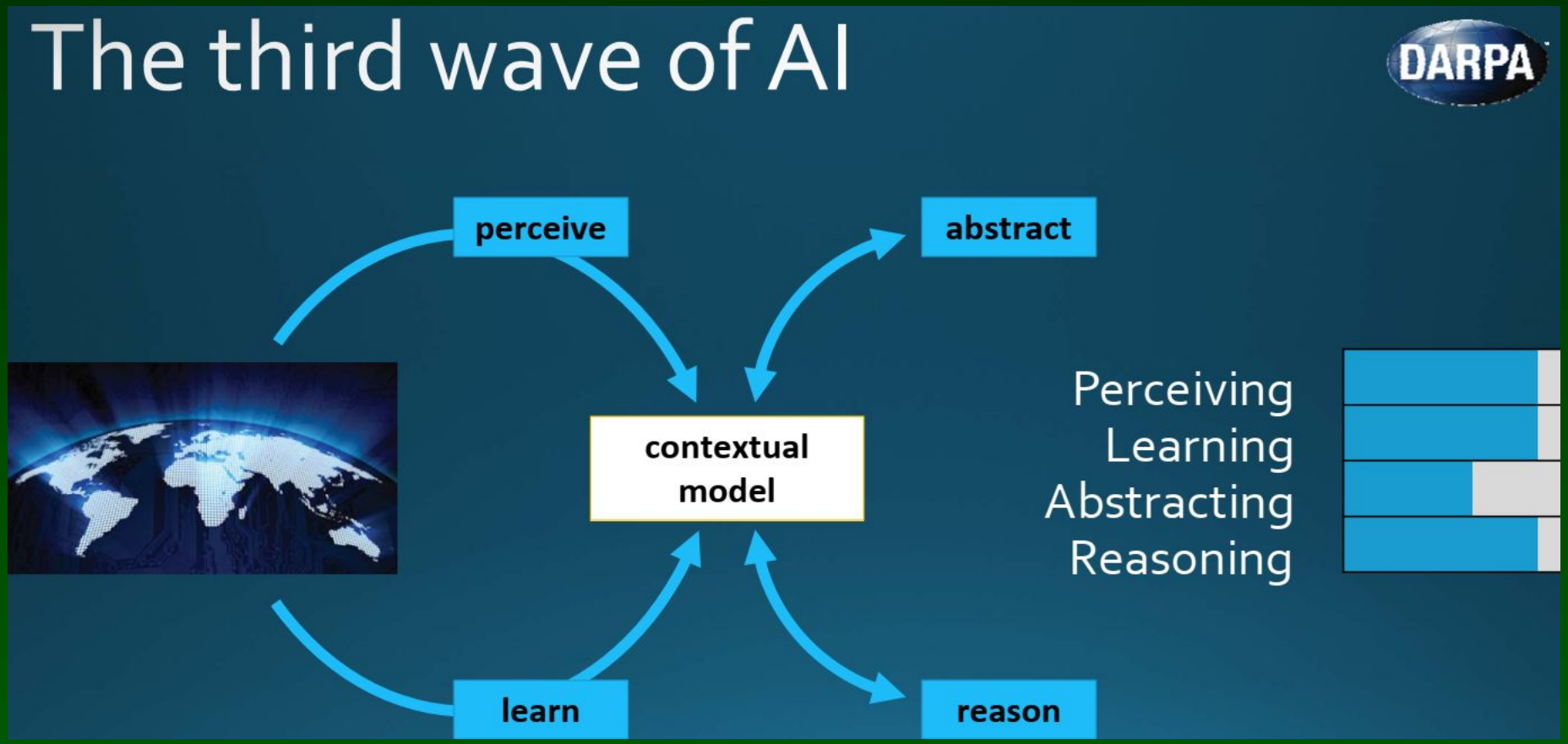
Włodzisław Duch
Katedra Informatyki Stosowanej UMK
Google: Włodzisław Duch

Co będzie

- Rozwój i zastosowania dużych modeli językowych

- Slajdy z referatów:
  State of Artificial Intelligence, Part I & II. 4/2023
- Referat dodatkowy, 1.06.2023
- Roadmap to Learn AI in 2024
  A free curriculum for hackers and programmers to learn AI

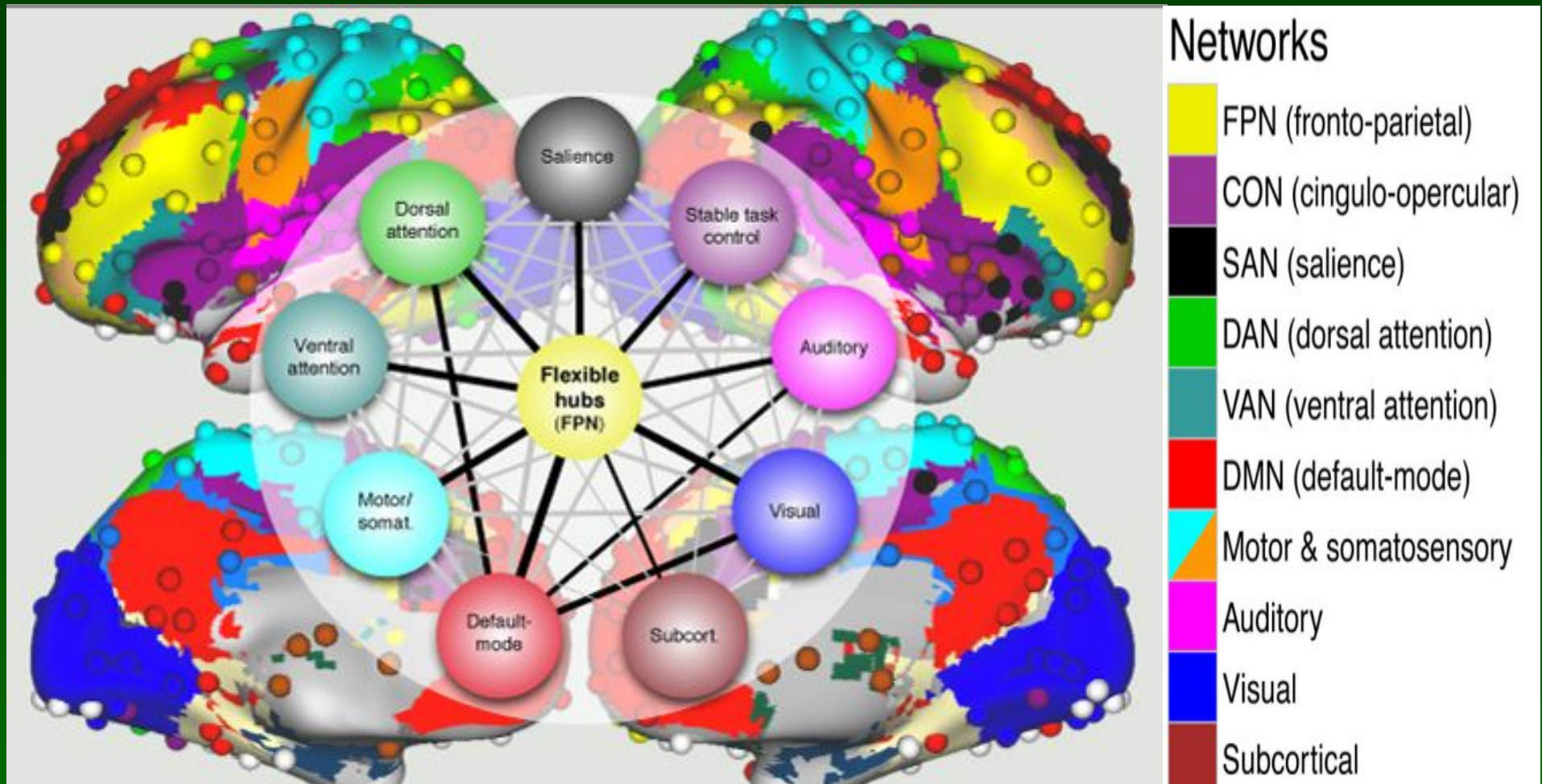
Third wave of AI



Final frontier: building models of objects and situations is the next step.
GAN, Generative Adversarial Networks, one network creates false examples distorting learning data, another network learns to distinguish them from natural ones.

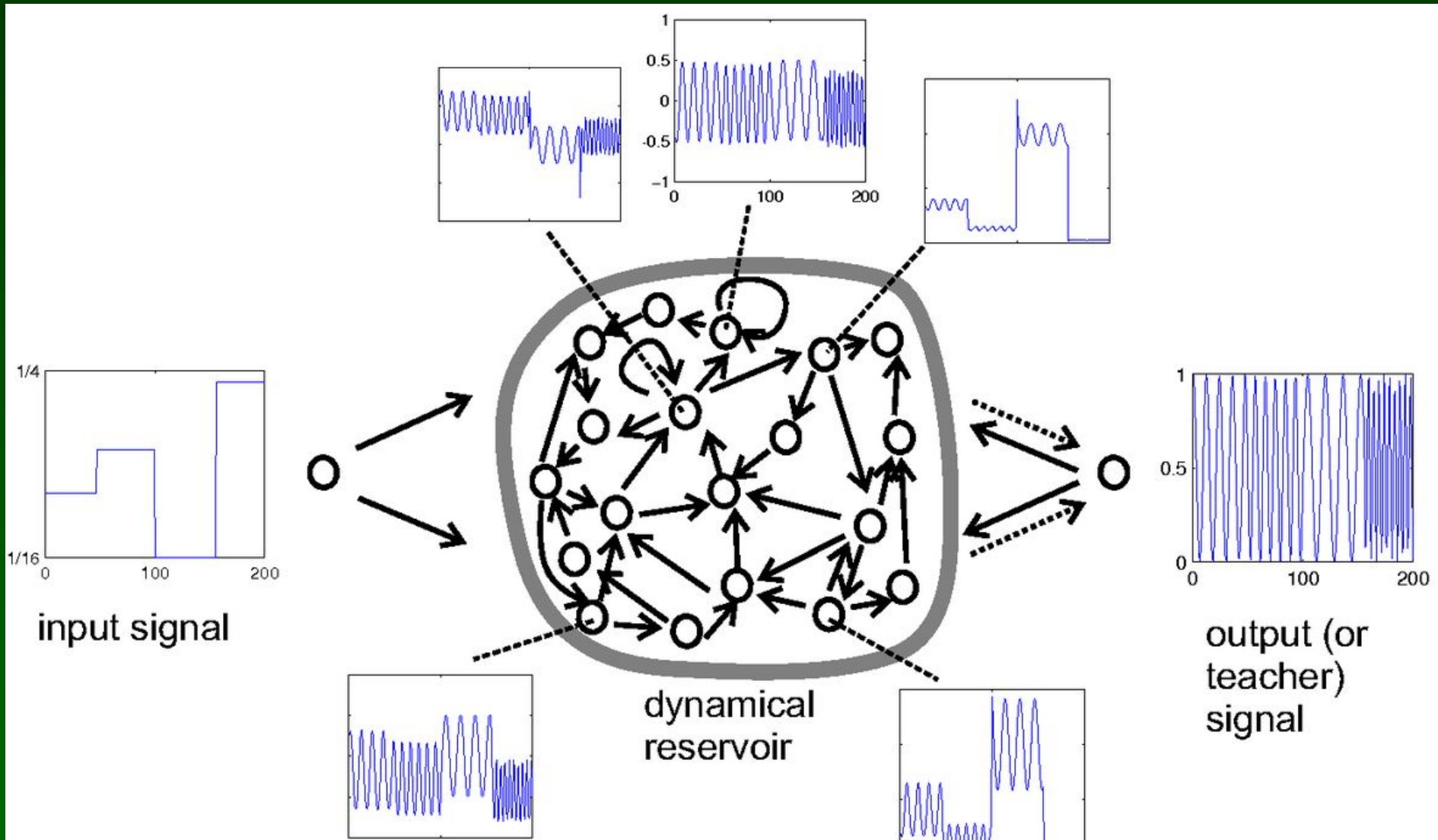
Brain activity ⇔ Mental image

fMRI activity can be correlated with deep CNN network features;
using these features closest image from large database is selected.
Horikawa, Kamitani, Generic decoding of seen and imagined objects using hierarchical visual features. Nature Comm. 2017.



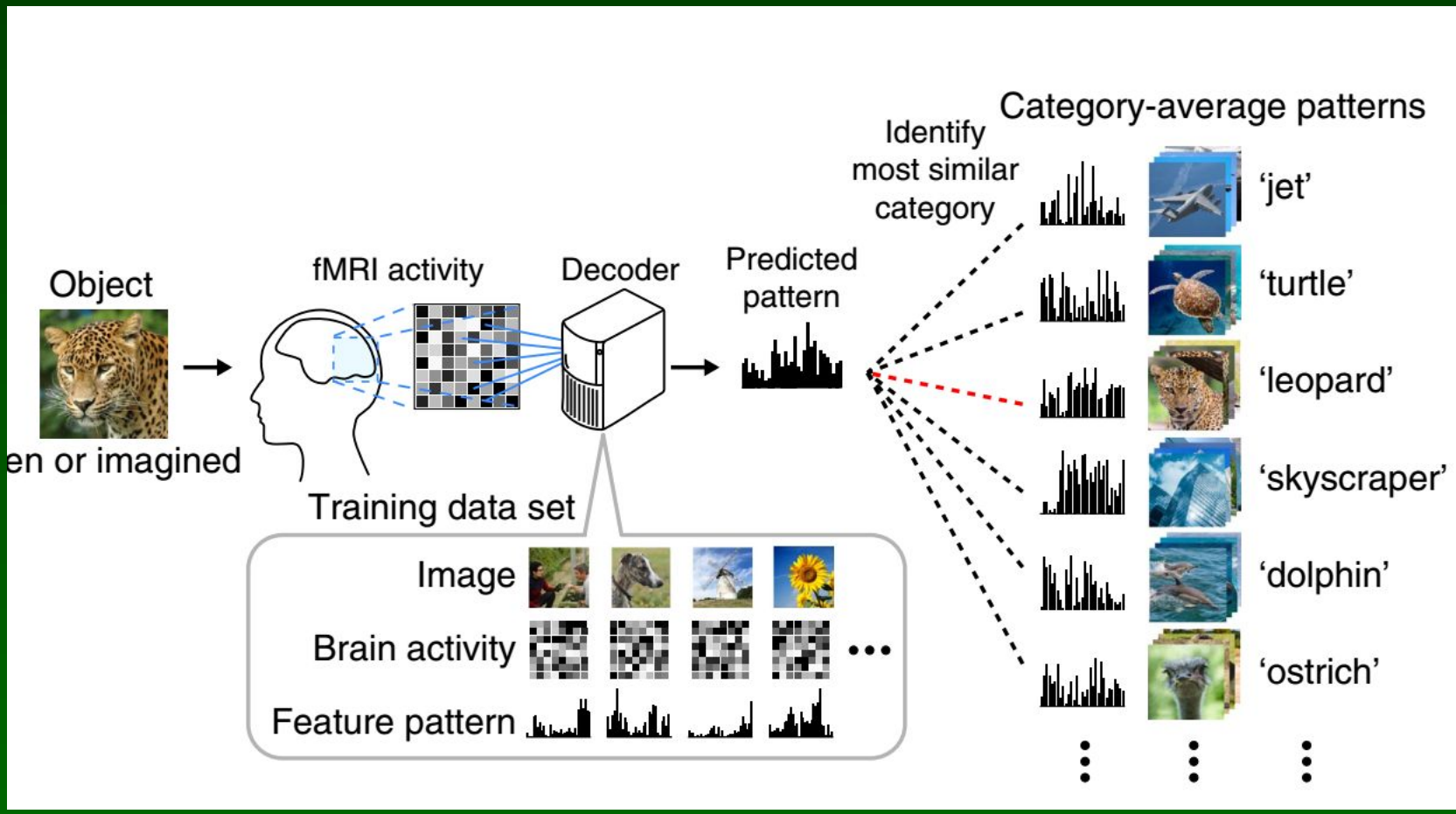
Brain activity ⇔ Mental image

fMRI activity can be correlated with deep CNN network features;
using these features closest image from large database is selected.



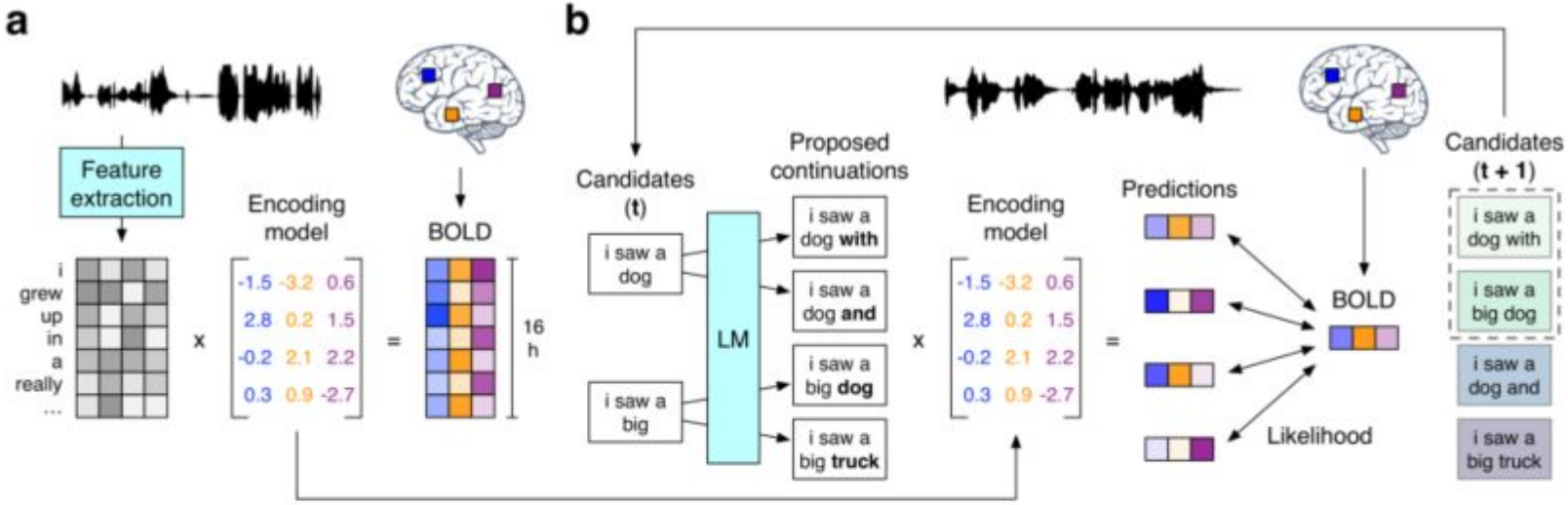
Brain activity ⇔ Mental image

fMRI activity can be correlated with deep CNN network features;
using these features closest image from large database is selected.
Horikawa, Kamitani, Generic decoding of seen and imagined objects using hierarchical visual features. Nature Comm. 2017.



Semantic reconstruction of continuous language

Tang, J., LeBel, A., Jain, S., & Huth, A. G. (2023). Semantic reconstruction of continuous language from non-invasive brain recordings. Nature Neuroscience, 26(5)



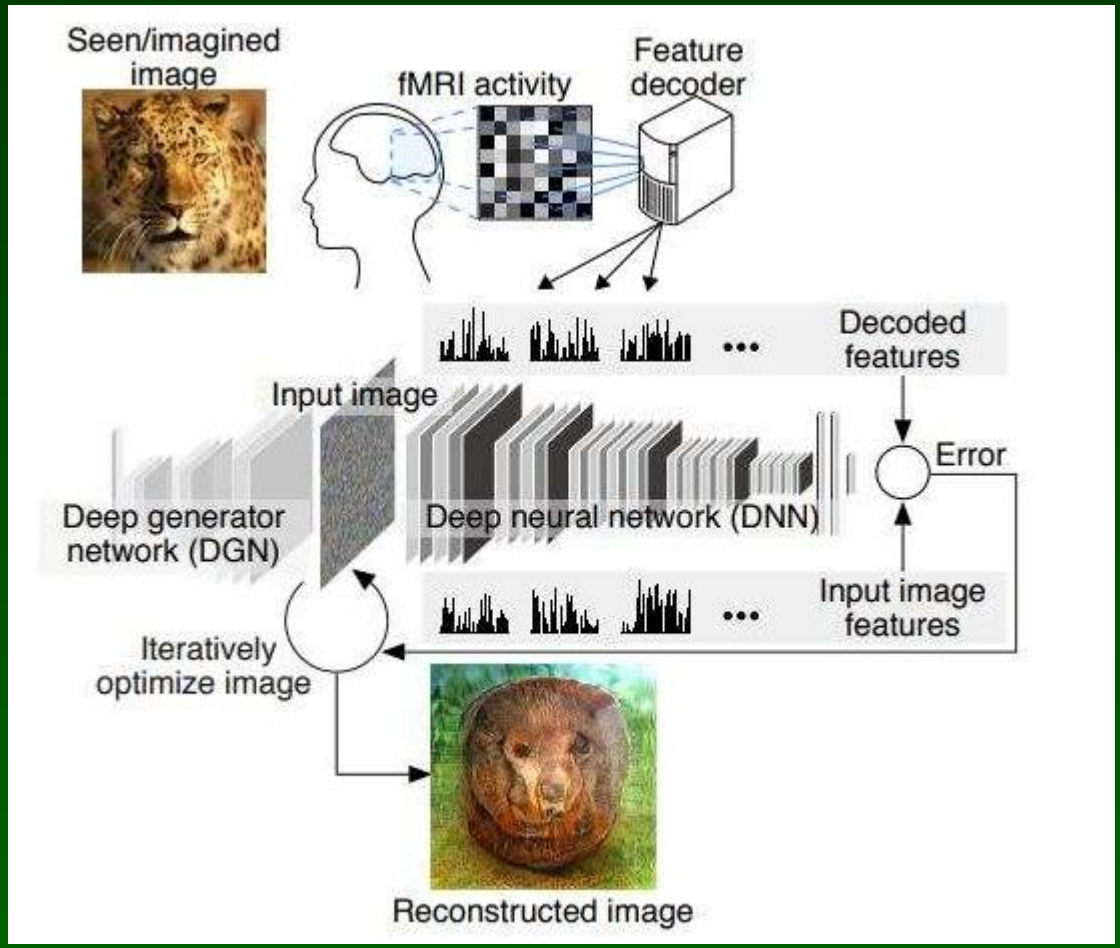
fMRI ⇔ CNN

Convert activity of the brain into the mental images that we are conscious of.

Try to estimate features at different brain areas/cortical layers.

8-layer convolution network, ~60 mln parameters, feature vectors from randomly selected 1000 units in each layer are used to represent images at different level of processing.

Output: vector of features that may be used to reconstruct image.



Interfejs mózg-robot

Australia, UTS: VR to control robotic dogs using EEG.
Dry graphene sensors, not as accurate as wet.
Przydatny?

Transformery

Więcej w moich prezentacjach z marca/maja 2024.

Attention is all you need (Transformer) - Model explanation (including math), Inference and Training

Wprowadzenie: Generative AI exists because of the transformer

RAG

In numerous instances, clients possess extensive proprietary documents, such as technical manuals, and require the extraction of specific information from this voluminous content. This task can be likened to locating a needle in a haystack.

Recently, OpenAI introduced a novel model, GPT4-Turbo, which boasts the capability to process large documents, potentially addressing this need. However, this model is not entirely efficient due to the “Lost In The Middle” phenomenon. This phenomenon mirrors the experience where, akin to reading the Bible in its entirety but struggling to recall what follows the Book of Samuel, the model tends to forget content located towards the middle of its contextual window.

To circumvent this limitation, an alternative approach known as Retrieval-Augmented-Generation (RAG) has been developed. This method involves creating an index for every paragraph in the document. When a query is made, the most pertinent paragraphs are swiftly identified and subsequently fed into a Large Language Model (LLM) like GPT4. This strategy of providing only select paragraphs, as opposed to the entire document, prevents information overload

Na zakończenie

Why are we confident that machines will pass Turing test …
—Rodney Brooks, director of the MIT AI Lab

There's this stupid myth out there that A.I. has failed, but A.I. is everywhere around you every second of the day. People just don't notice it. You've got A.I. systems in cars, tuning the parameters of the fuel injection systems. When you land in an airplane, your gate gets chosen by an A.I. scheduling system. Every time you use a piece of Microsoft software, you've got an A.I. system trying to figure out what you're doing, like writing a letter, and it does a pretty damned good job. Every time you see a movie with computer-generated characters, they're all little A.I. characters behaving as a group. Every time you play a video game, you're playing against an A.I. system.

Machine Learning and Creativity (wykład mniej techniczny).

AI Links wiele ciekawostek.

Na zakończenie

RAG, Retrival Augumented Generation
Chains of thought, step by step
Tree of toughts,
Książka o promptach

Halucynacje LLM, przyczyny, antidotum – faktt checking, kilka LLM,
Ranking: 3/24, Leaderboard Vectara,
GPT-4, GPT 4 Turbo 3.0 %
Microsoft Orca-2-13b 3.2 %
GPT 3.5 Turbo 3.5 %
Google Gemini Pro 4.8 %
Llama 2 7B, 13B, 70B 5.1-5/9 %
Anthropic Claude 3 Sonnet 6.0 % Opus 7.4 %

Przykładowe pytania

- Jakie mamy klasy metod uczenia maszynowego?
- Na czym polega ML?
- Przedstawić algorytm drzew decyzji, jakie granice tworzy?
- Etapy data mining
- Jaką metodę uczenia maszynowego warto stosować w sytuacji …