GraphNEx

Graph Neural Networks for Explainable Artificial Intelligence

Queen Mary University of London

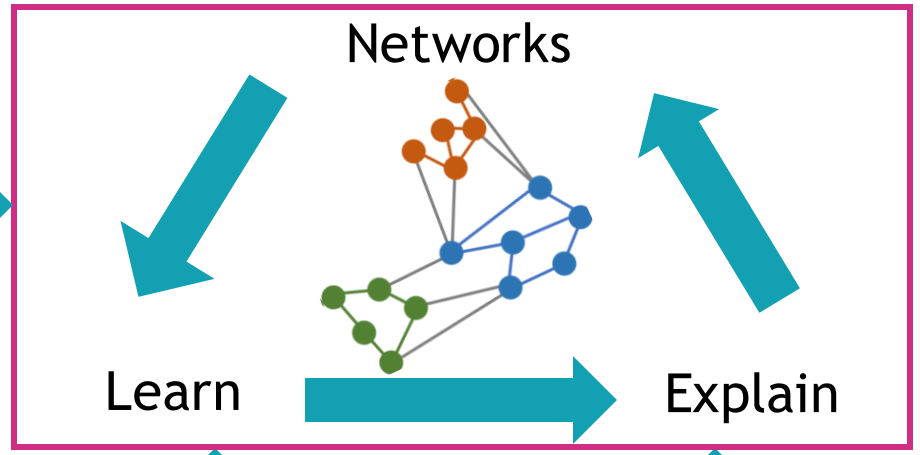
EPFL

Université de Lyon

ENS de Lyon

graphnex.eecs.qmul.ac.uk

Multiple sources of information



Semantic reasoning

- Graph construction & refinement
- Semantic concepts
- Rules discovery
- Attention-driven concept selection

Systems genetics

classic approaches
fail to capture patterns
across biological layers
(only pairwise comparisons)

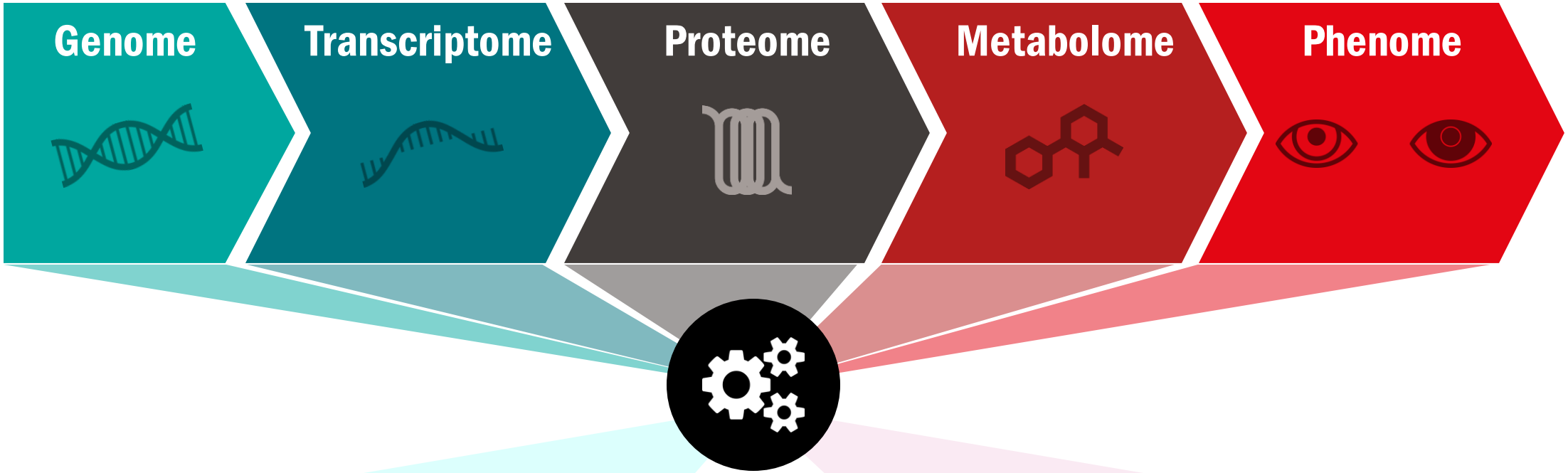
To capture complex patterns
across multiple biological data layers
& highlight key elements of these patterns

Privacy protection

secondary information of AI systems:
irrelevant for intended purpose
but may bias the prediction
& infringe upon users' privacy

To discover complex relationships between
similar, hierarchical or correlated (private)
concepts in audio and video data

GraphNEx
Graph Neural Networks
for Explainable Artificial Intelligence



Disease

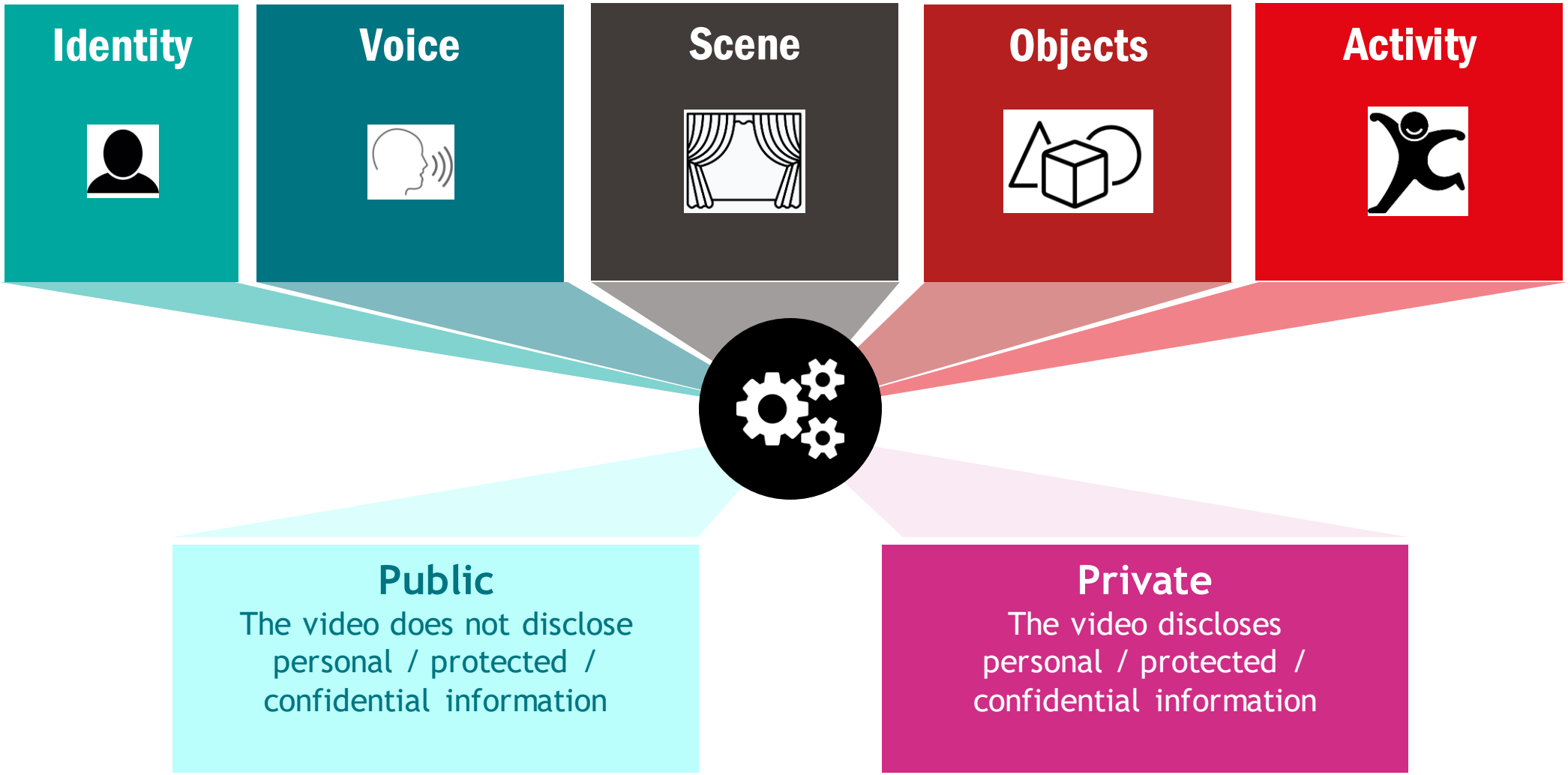
GeneY associated with disease
AND High expression in Liver
AND High blood pressure

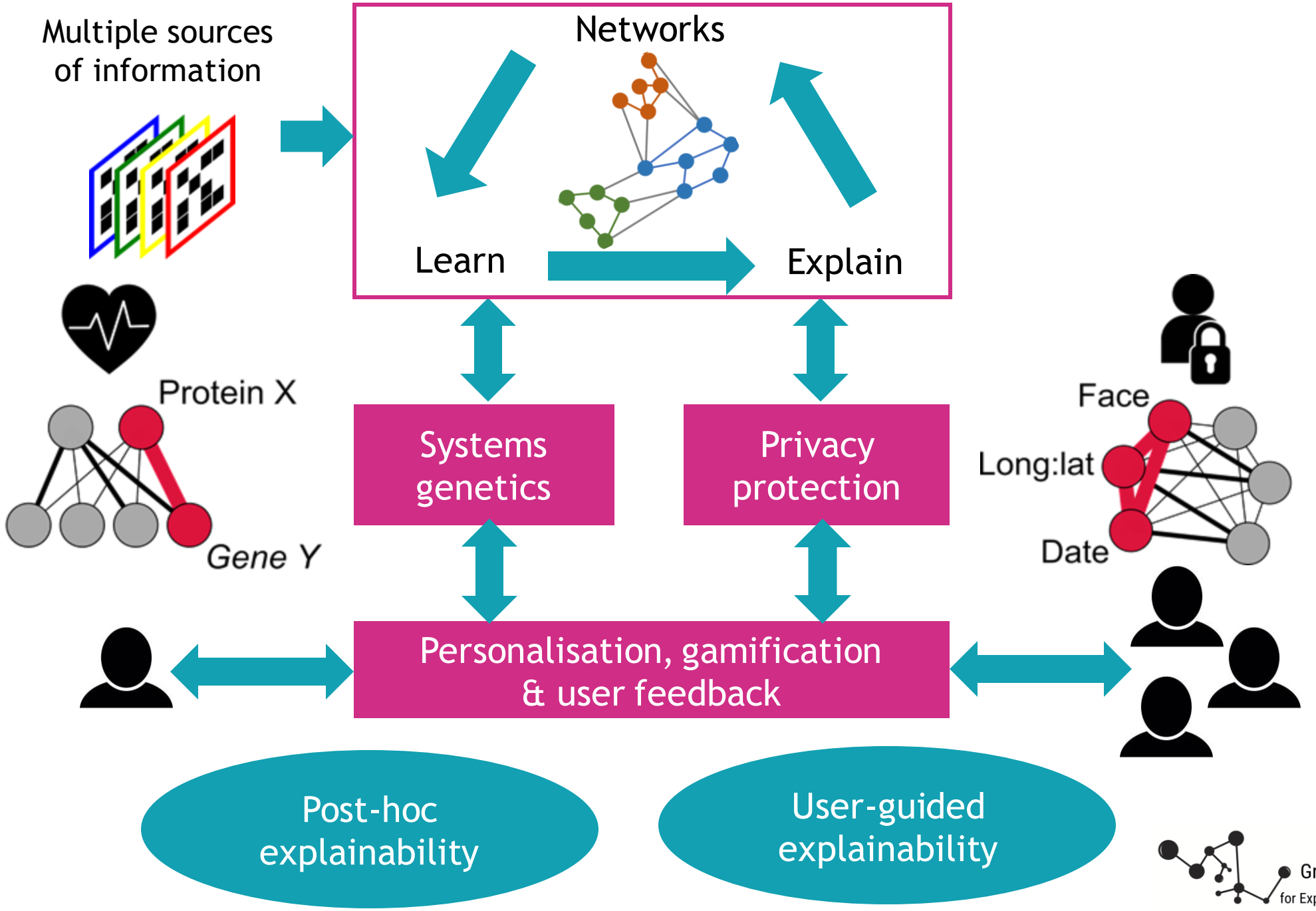
No disease

GeneX associated with no disease
AND Low expression in Liver
AND Normal blood pressure

Adapted from:

- Li, H. et *al.*, An Integrated Systems Genetics and Omics Toolkit to Probe Gene Function. *Cell Syst*, 2018
- Zitnik, M. et *al.*, Machine learning for integrating data in biology and medicine: Principles, practice, and opportunities. *Information Fusion*, 2019


Identity
Voice
Scene
Objects
Activity
Public
The video does not disclose personal / protected / confidential information
Private
The video discloses personal / protected / confidential information


Multiple sources of information
Networks
Learn
Explain
Protein X
Gene Y
Face
Long:lat
Date
Systems genetics
Privacy protection
Personalisation, gamification & user feedback
Post-hoc explainability
User-guided explainability
GraphNEx
Graph Neural Networks
for Explainable Artificial Intelligence

Phenotype prediction from gene expression data

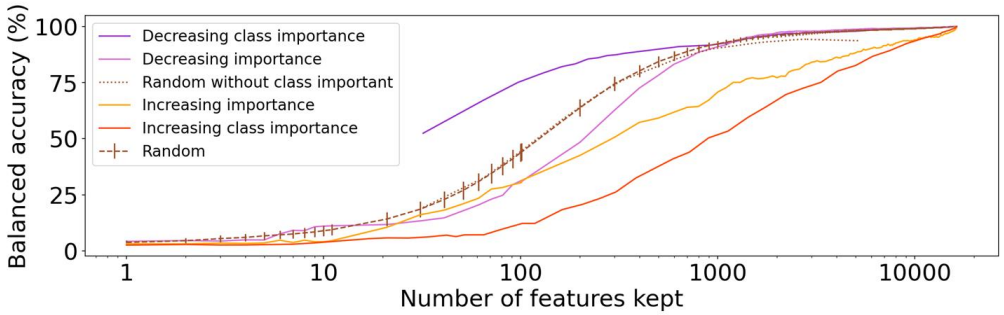
Good practices for evaluating explanations

Explainability metrics
Prediction gaps on important (PGI) and unimportant features (PGU)

Performance metric
Balanced accuracy (BA)

	Logistic regression	Multilayer perceptron
BA (↑)	93.2 %	94.7 %
PGI (↑)	0.957	0.957
PGU (↓)	0.003	0.020

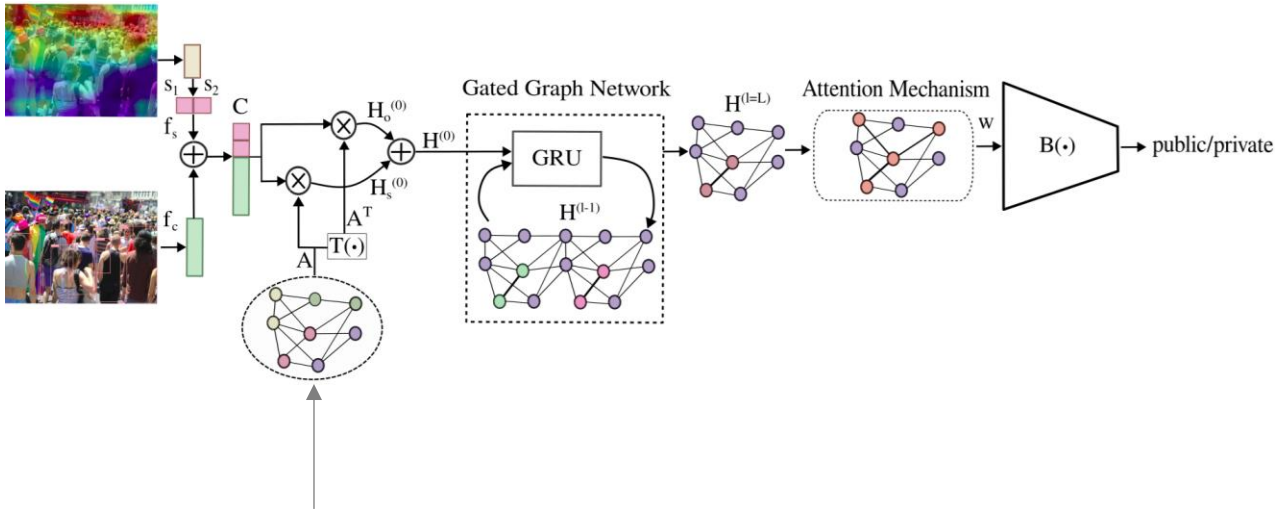
XAI: Feature attribution



Studying Limits of Explainability by Integrated Gradients for Gene Expression Models
M. Bontonou, A. Haget, M. Boulougouri, J. Arbona, B. Audit, P. Borgnat
https://arxiv.org/pdf/2303.11336v1.pdf (Under review, 2023)

Image privacy classification

Cardinality of different object types localized in an image

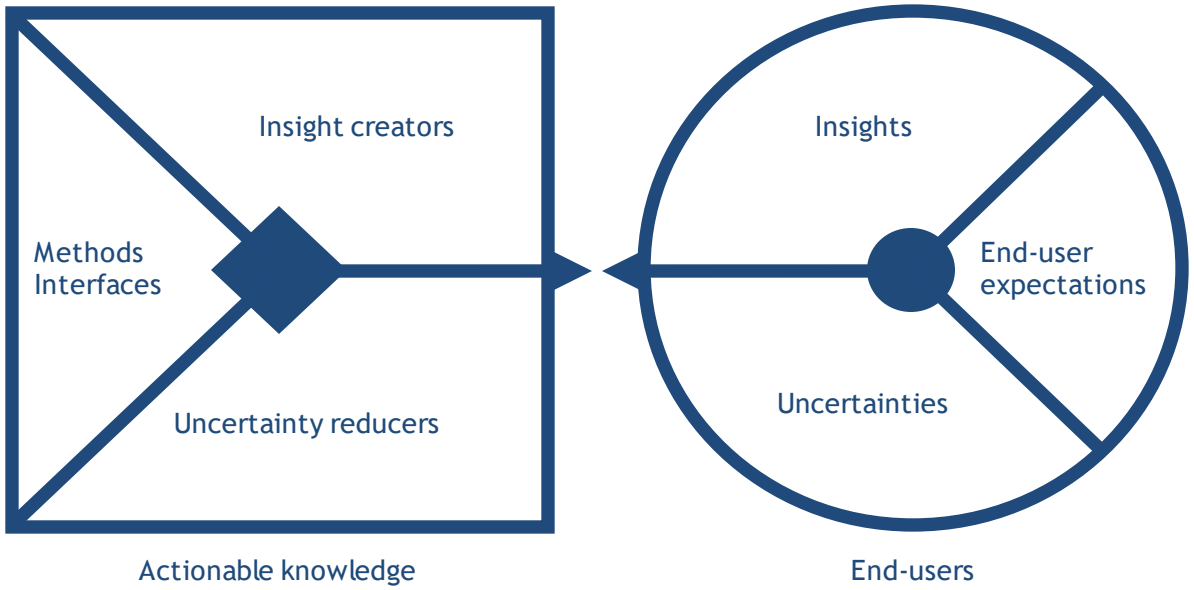


Prior knowledge graph with binary co-occurrence of objects

Content-based Graph Privacy Advisor
D. Stoidis, A. Cavallaro
IEEE International Conference on Multimedia Big Data (BigMM), December 2022
https://doi.org/10.48550/arXiv.2210.11169

Software and models:
https://github.com/smartcameras/GPA

Explainability Value Proposition Canvas (xVPC)



User-centered design of interfaces for explainability: privacy protection, system genetics, education

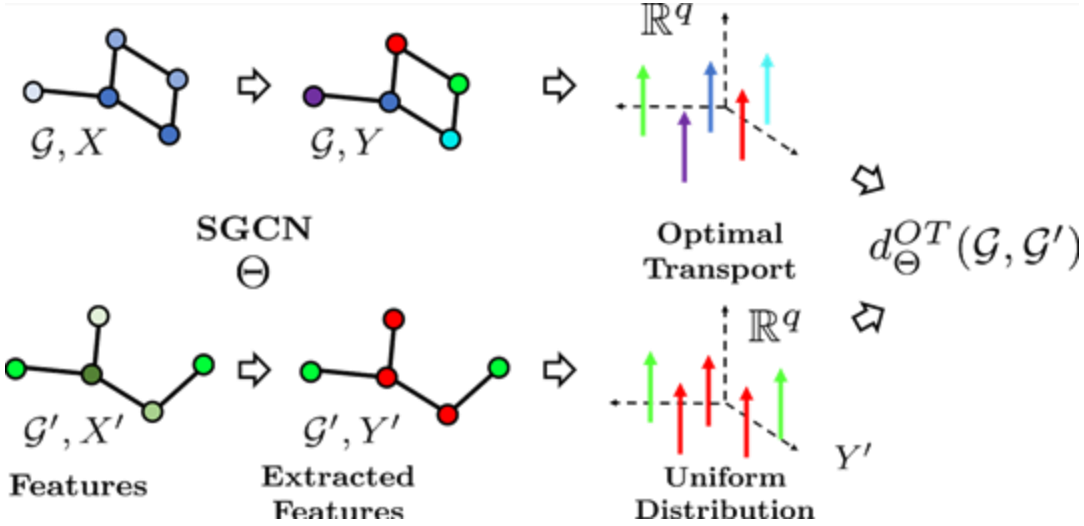
Introducing Alternative Value Proposition Canvases for Collaborative and Blended Design Thinking Activities in Science and Engineering Education
D. Gillet, I. Vonèche-Cardia, and J. La Scala
IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE), December 2022
https://infoscience.epfl.ch/record/297198

Privacy Awareness

Experiential dialogue between people and algorithms to stimulate the adoption of privacy-preserving behaviours

Reframing the narrative of privacy through system-thinking design
L. Ferrarello, R. Fiadeiro, R. Mazzon, A. Cavallaro
Design Research Society Conference (DRS), June-July 2022
https://doi.org/10.21606/drs.2022.620

Simple Graph Metric Learning Model (SGML)

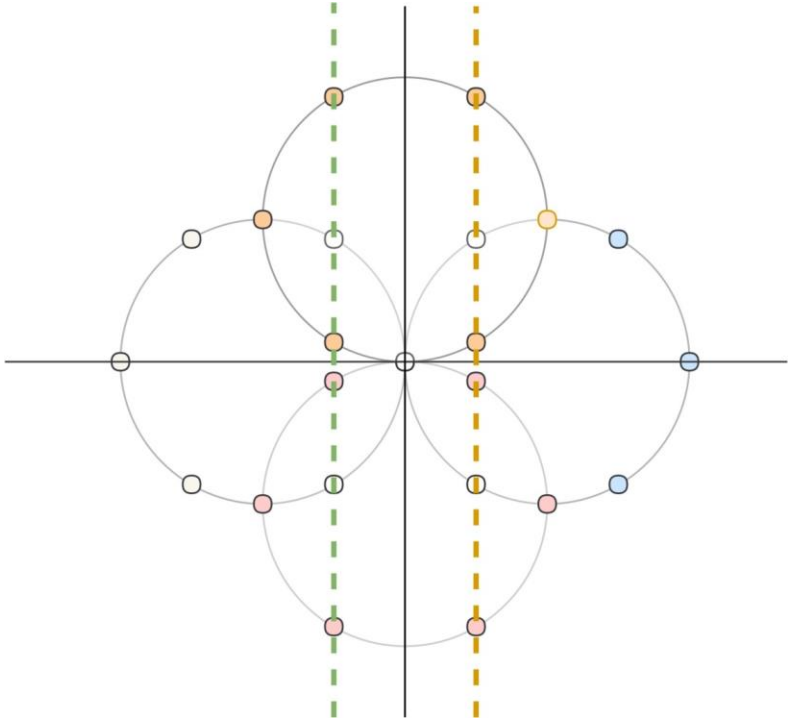


Elements of Optimal Transport theory
& few trainable parameters

Learn an appropriate distance to improve
the performance of simple classification algorithms

A Simple Way to Learn Metrics Between Attributed Graphs
Y. Kaloga, P. Borgnat, A. Habrard,
Learning on Graphs Conference (LoG), 2022
https://proceedings.mlr.press/v198/kaloga22a.html

Harmonic analysis on digraphs



Efficient mathematical & computational approach
for large and growing graph-data
with random walk operator

Harmonic analysis on directed graphs and applications: From Fourier analysis to wavelets
H. Sevi, G. Rilling, P. Borgnat
Journal of Applied and Computational Harmonic Analysis, 2023
https://doi.org/10.1016/j.acha.2022.10.003

Full list of publications at: graphnex.eecs.qmul.ac.uk/publications.html

2022 Intelligent Sensing Winter School, 12-14, 19 December

Explainable AI Sensing

4 Tutorials, 8 Talks

Speakers


Grégoire Montavon
Freie Universität Berlin
Germany


Ghassan AlRegib
Georgia Tech
USA


Gloria Menegaz
University of Verona
Italy


Myriam Bontonou
ENS Lyon
France


Vinod Subramanian
QMUL
UK


Wojciech Samek
Fraunhofer HHI
Germany


Denis Gillet
EPFL
Switzerland


Zhao Ren
Leibniz Universität Hannover
Germany


Yuki M. Asano
University of Amsterdam
The Netherlands


Seong Joon Oh
University of Tübingen
Germany


Stavros Ntalampiras
University of Milan
Italy


Srishti Gautam
UiT The Arctic University of Norway
Norway

Queen Mary University of London

http://cis.eecs.qmul.ac.uk/school2022.html

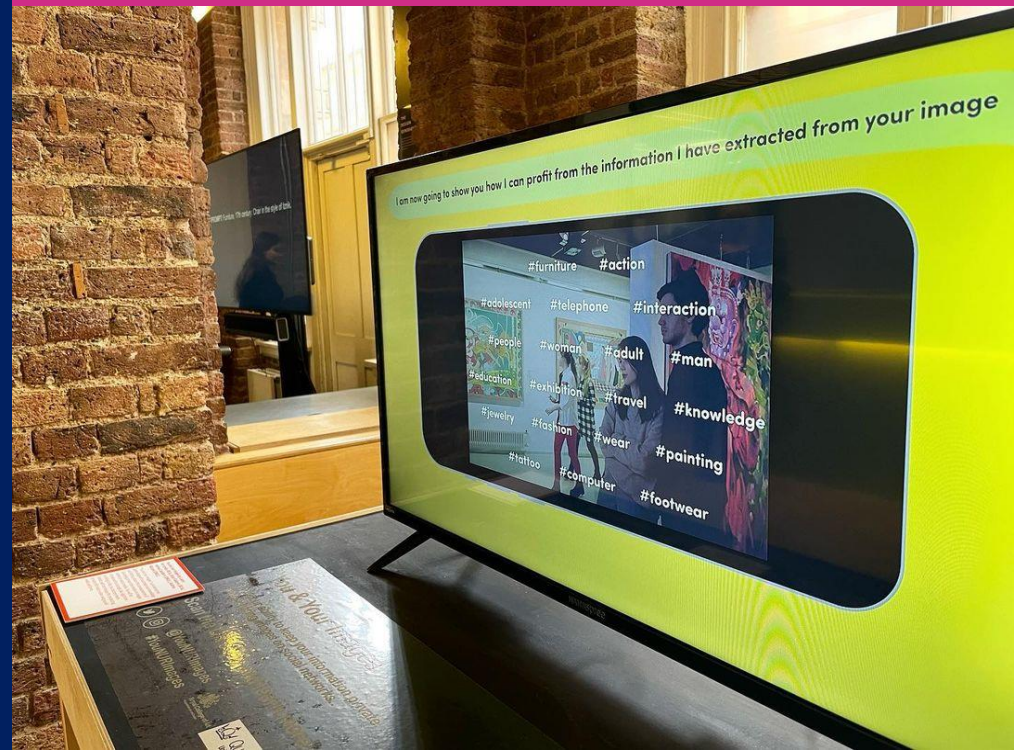
CIS centre for intelligent sensing

600+ registrations, 80-200 attendees/talk

GraphNEx talks:

- Using Explainable AI to Decipher Biological Rules
- GraphNEx XAI canvas

Victoria & Albert Museum (London, UK)
Friday Late event &
Digital Design Weekend 2022



2,000+ attendees at both events

100 participants to the "You and your images" demo activity for privacy awareness

GraphNEx

Graph Neural Networks for Explainable Artificial Intelligence

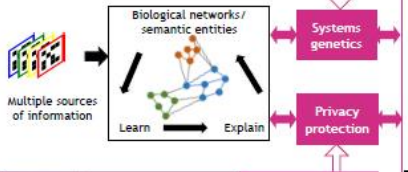
Objectives

- To combine semantic reasoning over knowledge bases with simple modular learning
- To extrapolate semantic concepts and meaningful relationships from sub-graphs (concepts)
- To enforce sparsity and domain-specific priors between concepts for human interpretability

Systems genetics

Studying Limits of Explainability by Integrated Gradients for Gene Expression Models
M. Bontonou, A. Haget, M. Boulougourt, J. Arbona, B. Audit, P. Borgnat

to assist the discovery of clinically and biologically relevant concepts on large, multimodal genetic and genomic datasets



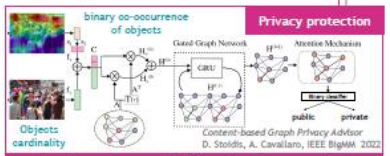
Interfaces for explainability

Personalisation Gamification User feedback

Introducing Alternative Value Proposition Canvases for Collaborative and Blended Design Thinking Activities in Science and Engineering Education
D. Gillet, I. Vonèche-Cardia, J. La Scala IEEE TALE 2022

Reframing the narrative of privacy through system-thinking design
L. Ferrarello, R. Fiadeiro, R. Mazzon, A. Cavallaro, DRS 2022

Privacy protection



Content-based Graph Privacy Advisor
D. Stoldis, A. Cavallaro, IEEE BigMM 2022

to safeguard personal information from unwanted non-essential inferences from multimedia data (images, video, audio) & support informed consent

Team

Project start: August 2021

Partners

Queen Mary University of London

ENS DE LYON

EPFL

GraphNEx Graph Neural Networks for Explainable Artificial Intelligence

graphnex.eecs.qmul.ac.uk

Sponsors

Engineering and Physical Sciences Research Council

anr agence nationale de la recherche

Swiss National Science Foundation

chist-era

chist-era

GraphNEx Graph Neural Networks for Explainable Artificial Intelligence

EPFL

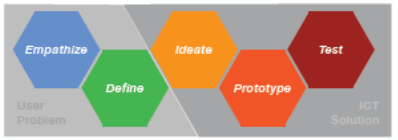
GraphNEx

Explainability Value Proposition Canvas

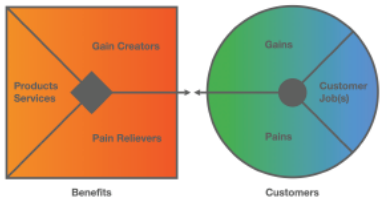
Denis Gillet, Basile Spaenlehauer, and Roxane Burri

Frontend	X	AI	Backend
User Experience (UX)		Machine Learning	
Human Computer Interaction		Graph Neural Networks (GNN)	
Interfaces designed for interacting with explanations		Methods designed for providing explanations	

User-centered design approaches needed



The value proposition canvas introduced in **business** can be used as a **boundary object** for **user-centered design** supporting the **Empathize**, **Define**, **Ideate**, and **Prototype** stages of **design thinking** activities (*Stanford Design School* model)



A Value Proposition Canvas has been proposed to support user-centered design of interfaces for explainability. It as been elicited through participatory design held during a GraphNEx workshop with experts from AI, human-computer interaction, and XAI application domains

Applied in GraphNEx for privacy protection*, system genetics, education in computer and data sciences

* Help average citizen understand what they disclosed in terms of private information when sharing pictures in social networks

Sources	Outputs
Data & Meta data Training data sets	Explanations Graphical and textual
Settings	**Analytics**
Preferences Links for comparison	History Comparisons

Core principles
Parsimony
Awareness
Personalization
Context
Reproducibility

Key user interface components for explainability

Denis Gillet, Isabelle Vonèche Cardia, and Jérémy La Scala,
Swiss Federal Institute of Technology in Lausanne (EPFL)

Introducing Alternative Value Proposition Canvases for Collaborative and Blended Design Thinking Activities in Science and Engineering Education, IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE), Hong Kong, December 4-7, 2022

Studying Limits of Explainability by Integrated Gradients for Gene Expression Models

Myriam BONTONOU[1], Anaïs HAGET[2], Maria BOULOUGOURI[2], Jean-Michel ARBONA[1], Benjamin AUDIT[3], Pierre BORGNAT[3]

[1]Univ Lyon, ENSL, CNRS, LBMC, Lyon, France [2]LTS2 laboratory, EPFL, Lausanne, Switzerland
[3]Univ Lyon, ENS de Lyon, CNRS, Laboratoire de physique, Lyon, France

ENS DE LYON

EPFL

GraphNEx Graph Neural Networks for Explainable Artificial Intelligence

Context

- **Supervised learning problems** are formulated to **decipher** complex **molecular processes** driving cellular life.
  E.g. phenotype prediction from transcriptomic data (gene expression).
- Feature attribution **explainability methods** return the **input features** on which the **individual predictions** are predominantly based.
- These features are often **interpreted as the cause of the phenotype**.

Problematic: What is the relevance of biomarkers identified using explainability methods?

GitHub

Contributions

- Exploration of the relevance of the features identified by explainability.
- Definition of **quantitative metrics**.
- **Simulation of data**, with known discriminative features, mimicking genes.

PyTorch code https://github.com/mbonto/XAI_for_genomics.

Definition of quantitative metrics

Sample level [2]
How the prediction of a sample changes when features are set to zero?
- Network f, input x, modified input $\tilde{x}$.

Prediction gap $\mathrm{PG} = \max(f(x) - f(\tilde{x}), 0)$

- **A feature under PG** when an increasing number of features is set to zero with
  - most important removed first → **PG on Important features (PGI)**.
  - less important removed first → **PG on Unimportant features (PGU)**.

Model level
How the accuracy of a network changes when genes are set to zero?
- Accuracy obtained with the most important features for the whole dataset.
- Accuracy obtained with random features.

Do known discriminative features stand out among the identified features?
- Number of relevant features $\mathcal{F}$ among the identified features $\mathcal{M}$.

Feature Attribution FA $= \frac{|\mathcal{F} \cap \mathcal{M}|}{|\mathcal{F}|}$

Simulation of gene expression data

Generative probabilistic model called Latent Dirichlet Association [3].
→ Known for document generation.

Individual samples (documents) are generated with a **fixed number** N **of sequencing reads** (words) associated with **metabolic pathways** (subjects).
- Prior η_p, proportion of genes expressed in pathway p.
- Prior α_c, proportion of pathways expressed in class c.
- Proportion of reads appearing in a pathway $\beta_p \sim \mathrm{Dirichlet}(\eta_p)$.

Generation of a sample s with N reads
Step 1 Draw the proportion of pathways $\theta_s \sim \mathrm{Dirichlet}(\alpha_c)$.
Step 2 For each read i,
- pathway assignment $p_i \sim \mathrm{Multinomial}(\theta_s)$,
- drawn gene $g_i \sim \mathrm{Multinomial}(\beta_{p_i})$.

Experimental setting

- **Simulated data** (9900) or Gene expression from **PanCan TCGA** (9680).
- **Classification problem** 33 classes.
- **Algorithm** Logistic Regression (LR), Multilayer Perceptron (MLP), Diffusion layer on a correlation graph (D).
- **Explainability method** Integrated Gradients (IG).

PanCan TCGA [1] - 16335 genes.
SIMU1/2 - 15000 genes. 1500 non-overlapping / 3000 overlapping pathways.

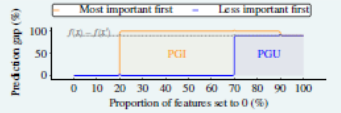


Figure 1 - Scheme describing the Prediction Gaps on Important features (PGI) and on Unimportant features (PGU).

Results

Table 1 - Explainability metrics averaged over test samples.

(a) Pan-Can TCGA

Network	LR	MLP	D + LR	D + MLP
Balanced accuracy (↑)	93.2%	94.7%	92.5%	94.3%
PGI (↑)	0.9570	0.9567	0.9750	0.9652
PGU (↓)	0.0035	0.0197	0.0053	0.0133

(b) Simulations

Dataset	SIMU1		SIMU2	
Network	LR	MLP	LR	MLP
Accuracy (↑)	99.5%	99.5%	99.9%	100%
PGI (↑)	0.9905	0.9714	0.9881	0.9842
PGU (↓)	0.0007	0.0036	0.0007	0.0039
FA (↑)	0.72	0.76	0.43	0.45
D + FA (↑)	1	1	0.96	1

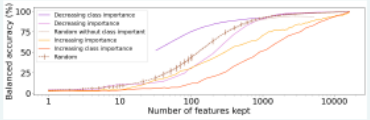


Figure 2 - Explainability metrics on Pan-Can TCGA data with LR.

Conclusion

- Evaluation of the **complexity of the real dataset PanCan TCGA.**
  - Set of 50 genes sufficient to classify each sample (PGU).
  - But not necessary (PGI).
- **Analyse** of the pertinence of the selected features **on simulated data** (FA).
- **Well behaved explanatory features are ambiguous.**

[1] https://portal.gdc.cancer.gov/.
[2] C. Agarwal, S. Krishna, E. Saxena, M. Pawelczyk, N. Johnson, I. Puri, M. Zitnik, and H. Lakkaraju. OpenXAI: Towards a transparent evaluation of model explanations. In *NeurIPS Datasets and Benchmarks Track*, 2022.
[3] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *JMLR*, 2003.

Come and see our three posters!

GraphNEx

Graph Neural Networks for Explainable Artificial Intelligence

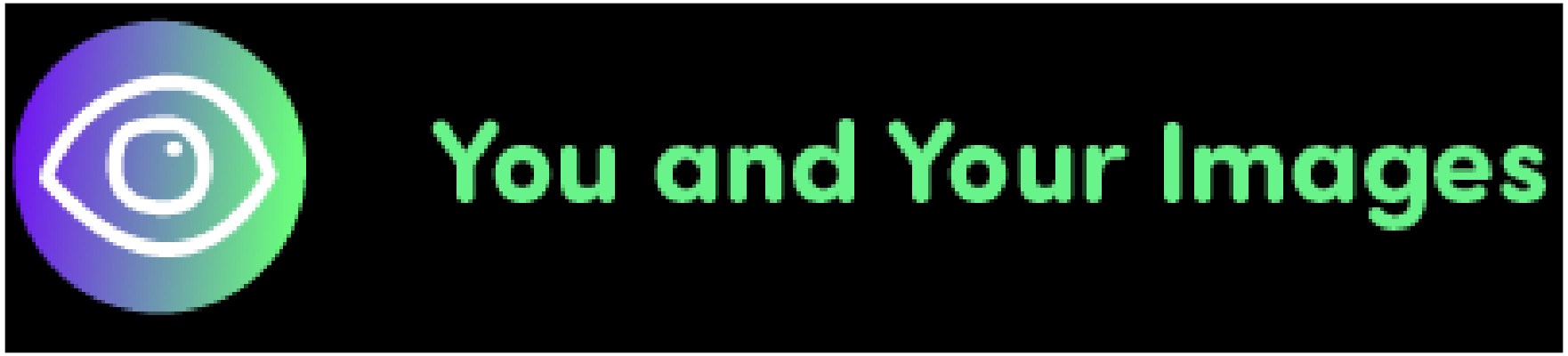
Queen Mary University of London

EPFL

Université de Lyon

ENS de Lyon

graphnex.eecs.qmul.ac.uk



Watch the video demo:
https://youtu.be/LR9Q_KTAhOs