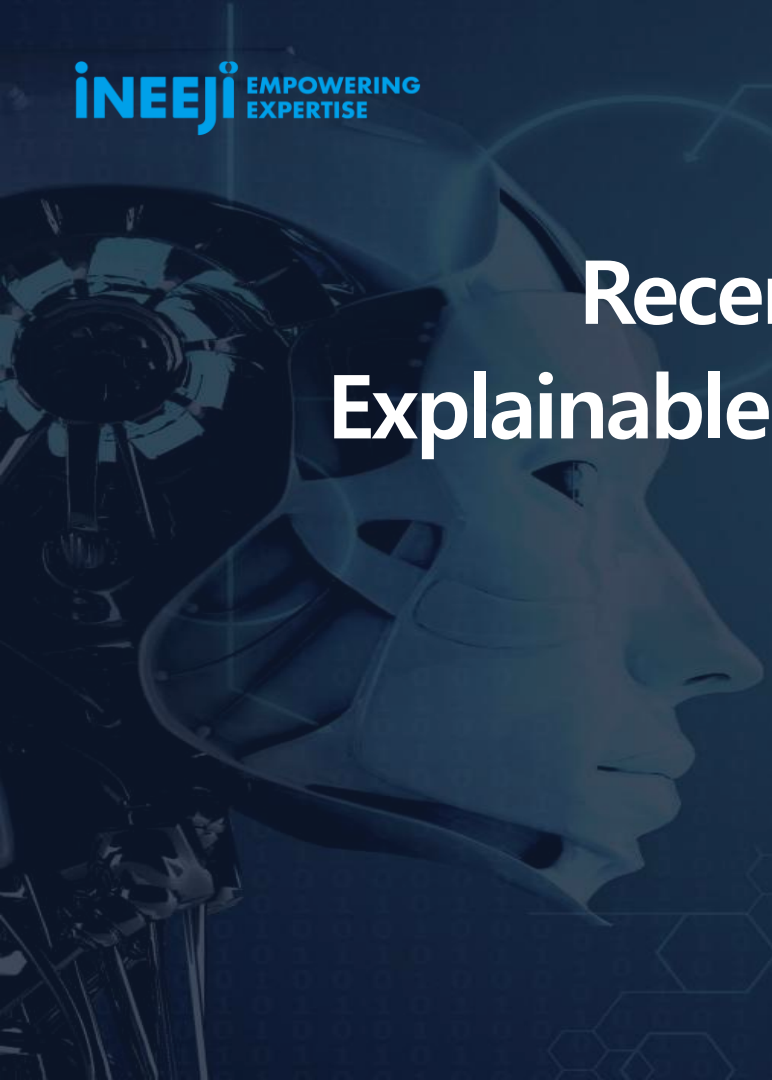
INEEJI EMPOWERING EXPERTISE

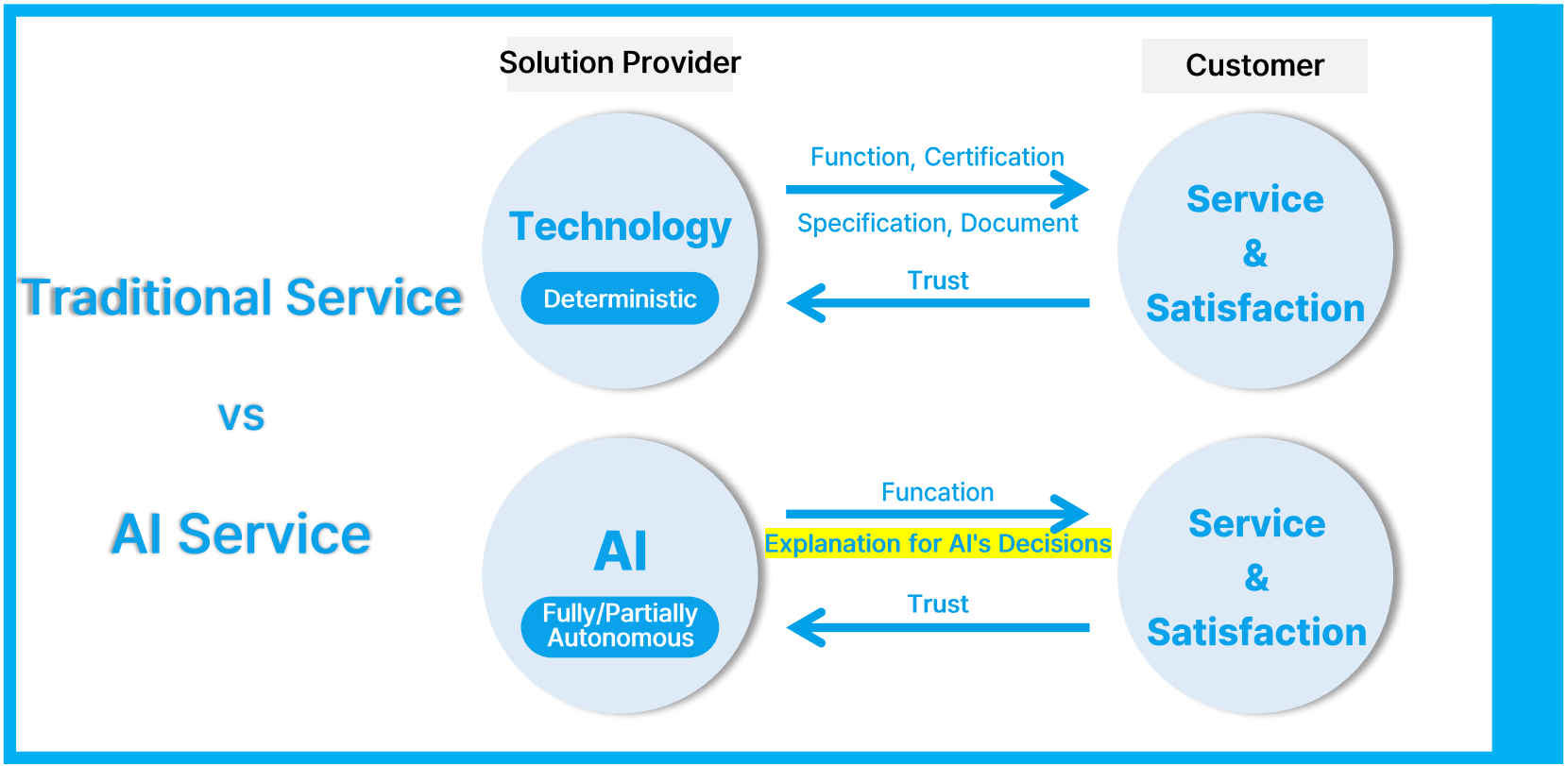


Recent Advances in Explainable Artificial Intelligence

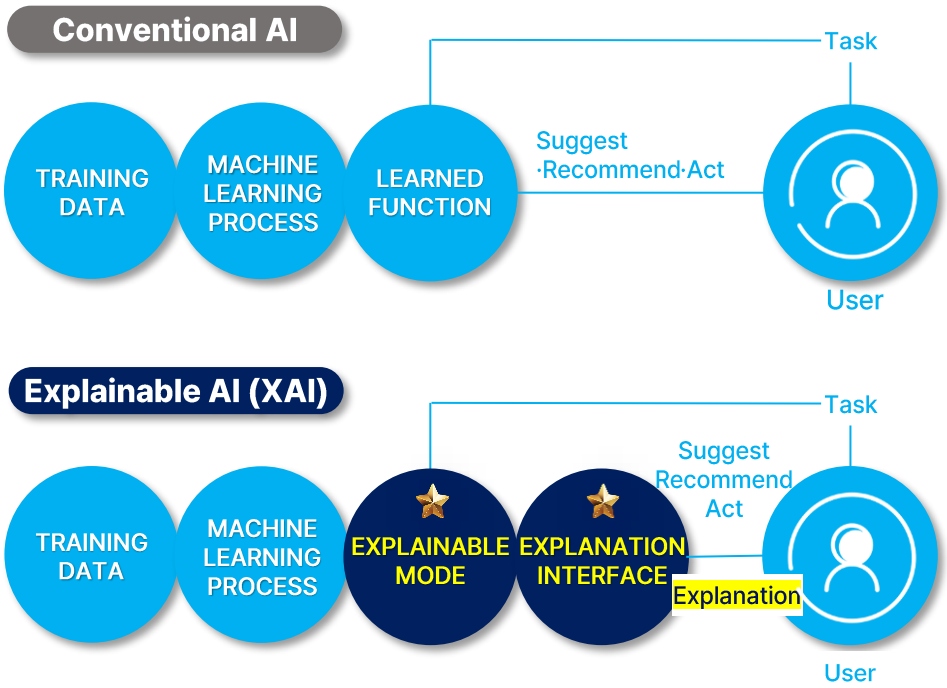
Professor, Kim Jaechul Graduate School of AI, KAIST
CEO, Ineeji Corp.

Jaesik Choi

Explainable AI (eXplainable AI)



Explainable AI (eXplainable AI)



- Why this conclusion?
- How did it succeed?
- How did it fail?
- Can it be trusted?
- How to fix?

- Conclusion is explained.
- Success is explained.
- Failure is explained.
- We can trust it in this case.
- You know why it occurred.

[Source: Defense Advanced Research Projects Agency (DARPA) under the U.S. Department of Defense]

Explainable AI in Korea

Trustworthy AI in Korea

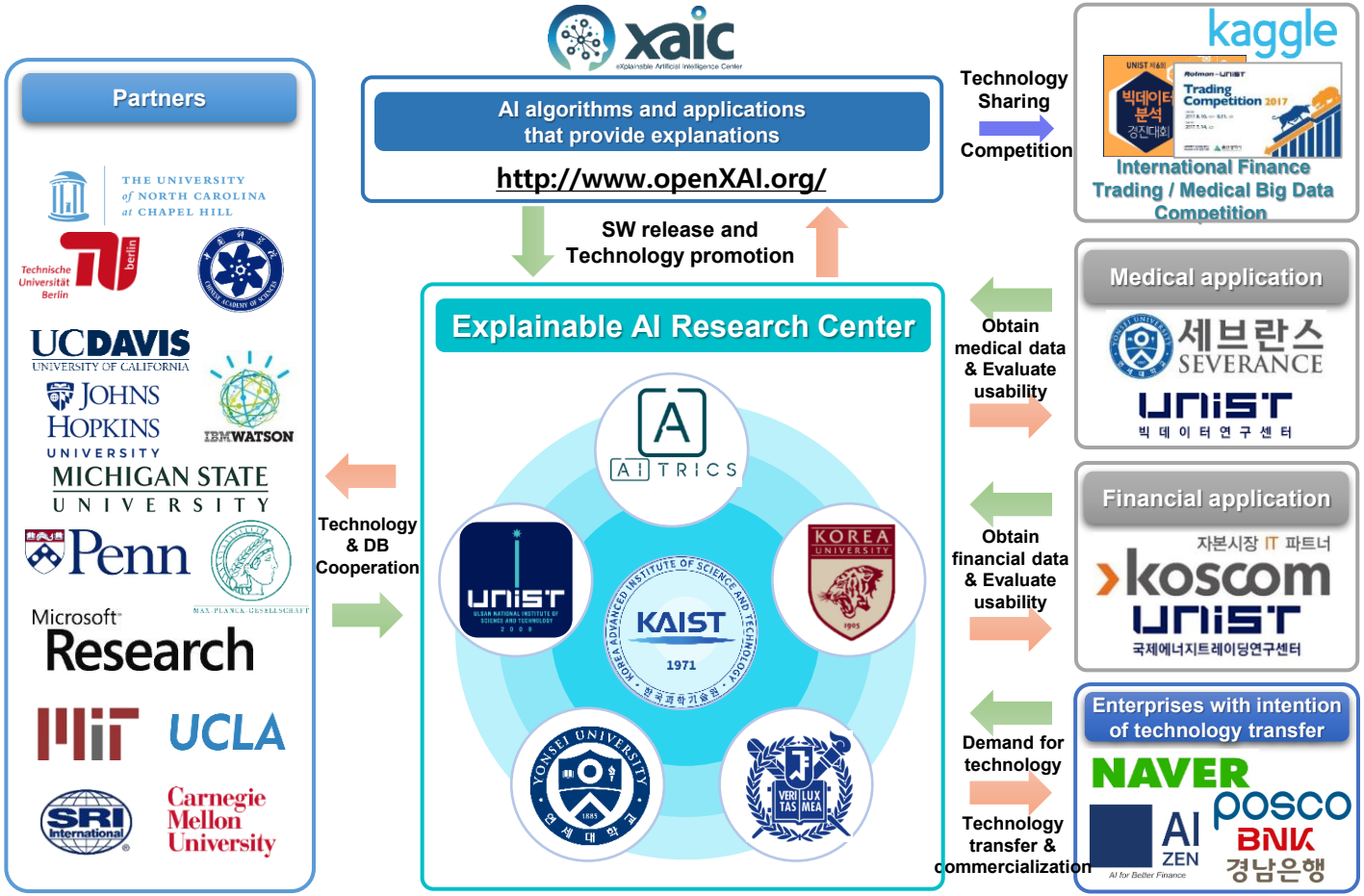
The strategy has the vision of "realize trustworthy artificial intelligence for everyone" and will be implemented step by step until 2025, based on the three pillars of 'technology, system, ethics' and 10 action plans.

Vision, goals, detailed strategies of trustworthy artificial intelligence

Vision	**Trustworthy AI for Everyone**		
Goal (~2025)	**Responsible use of AI** Global no.5	**Trustworthy society** Global no.10	**Safe cyber nation** Global no.3
Strategies	**Create an environment for trustworthy AI**	**Lay the foundation for safe use of AI**	**Spread AI ethics across society**
	1) Put in olace a systematic process for securing trust for AI products and services 2) Support players in the private sector with securing trust for AI 3) Developing source technology for trustworthy AI	1) Make AI learning data more trustworthy 2) Promote securing trsut for high-risk AI 3) Conduct assessment on influence of AI 4) Improve regulations for increased trust for AI	1) Provide strengthened education programs on AI ethics 2) Create and distribute checklists for each stakeholder 3) Operate a platform for policies on ethics

Explainable AI Program in Korea – Part I

July 2017 ~ December 2021 (54 months)

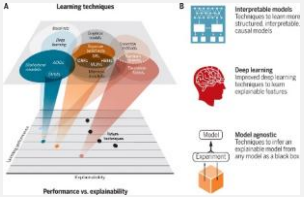


Explainable AI Program in Korea – Part I

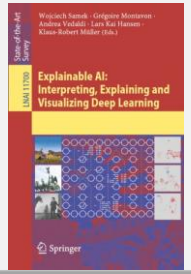
Research Results

AI Top Conference Papers (ICML, NeurIPS, AAAI, …) **87**

Top Journal Papers (Science Robotics, …) **45**



Book Edited (Explainable AI: Springer)



Technology Transfer

Patents (Registration) **37(2)**

Industrial Projects **11**

Manufacturing	Healthcare	Finance	Mobile
Process Explain	ICU monitoring	Credit Rating	Robust Generation

Open Source/Meetings

Open Source Projects github.com/OpenXAIProject **44**

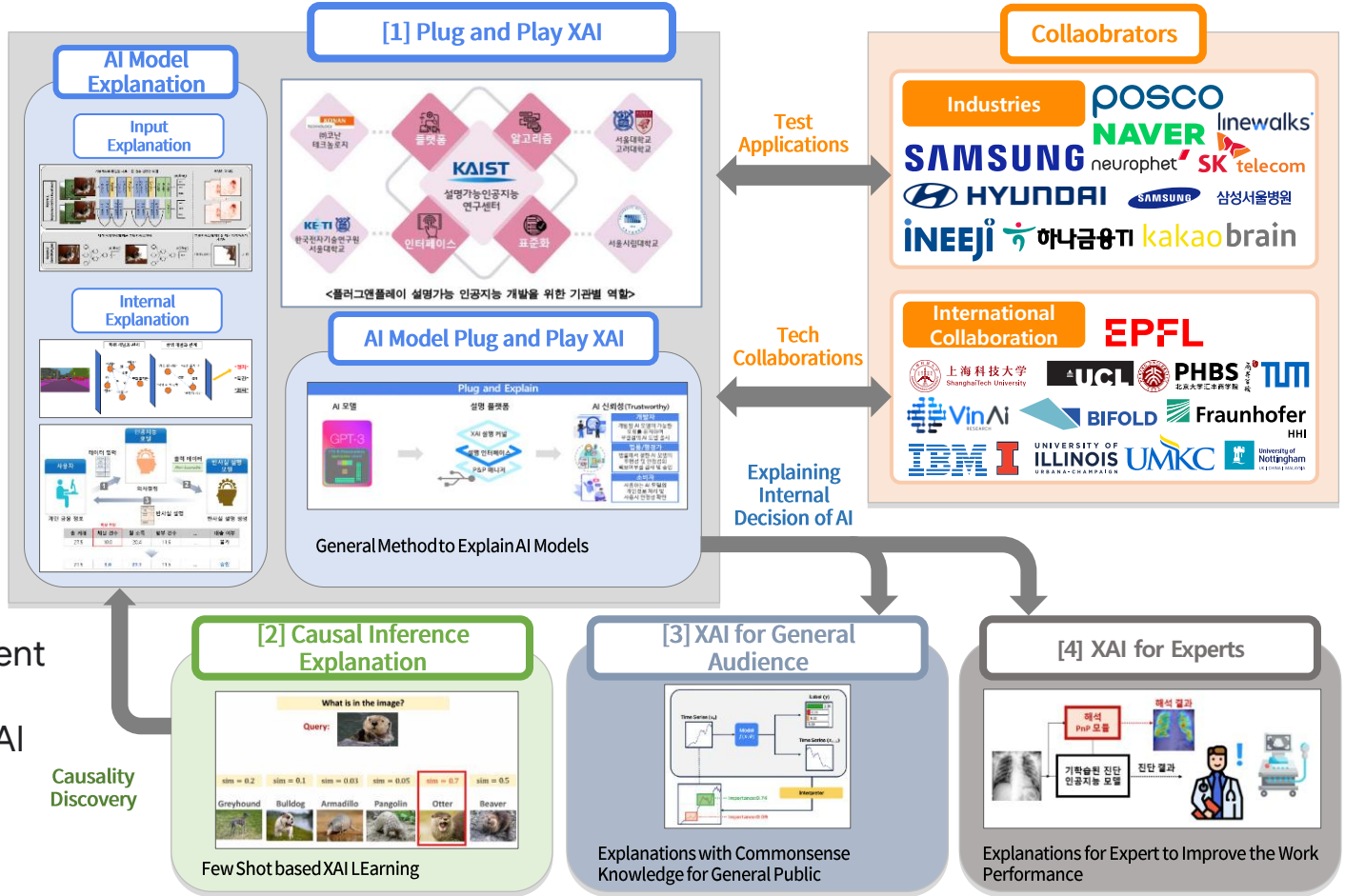
Online Tutorial **31**

Open Workshop **10**

International Gathering **3**

Explainable AI Program in Korea – Part II

April 2022 ~ December 2026 (57 months)



Korean Government Invest 45.0B KRW (32.6M USD) on XAI

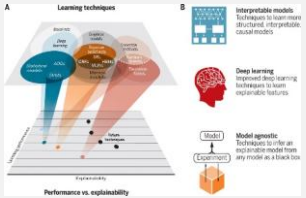
EMPOWERING EXPERTISE

Achievements

International Research Achievements

Technical papers at world's top-tier conferences in AI (e.g., ICML, NeurIPS, AAAI) — 106 papers

Papers in top-ranked journals (IF 4.0 or higher) — 54 papers



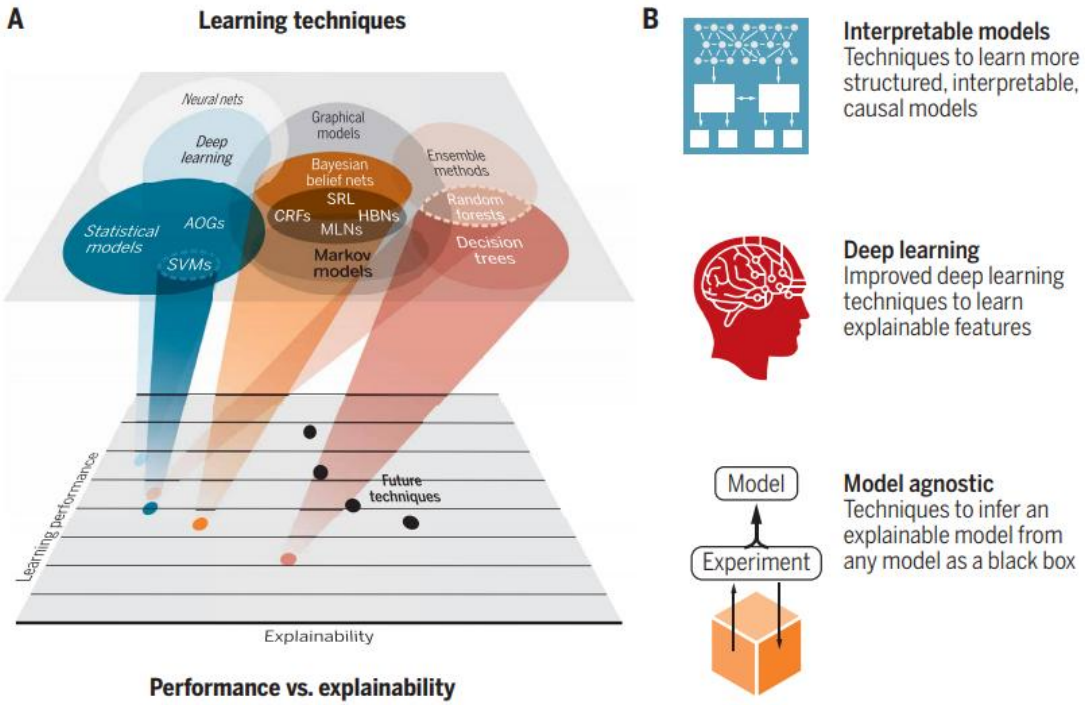
Intellectual Property, Copyright, and Technology Dissemination

Patent applications	51 applications
Software registration with the Korea Copyright Commission	2 registrations
Technology briefings/tutorials held	12 sessions

Technology Disclosure and Leadership in International Standards

Software releases	57 releases
Led the world's first international standard for XAI Standard proposal adopted ISO/IEC JTC 1/SC 42 (AI)	18 releases

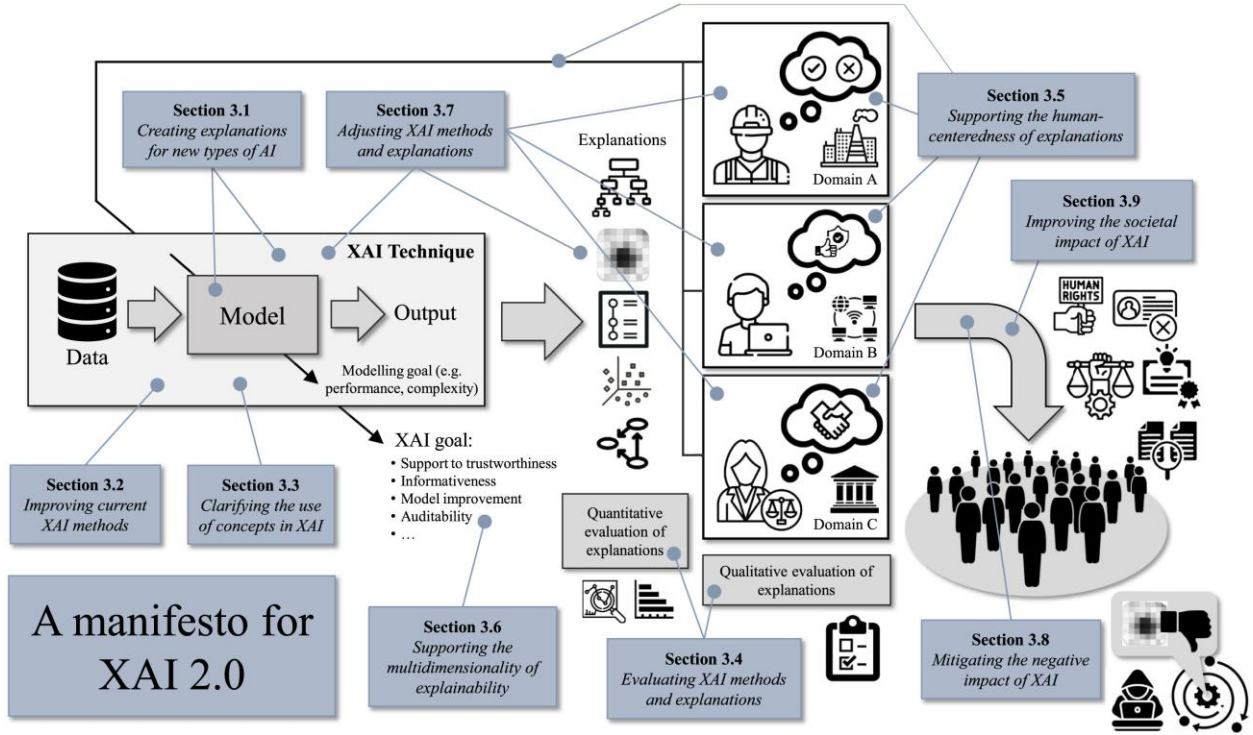
XAI - Explainable Artificial Intelligence



Performance vs. explainability

David Gunning, Mark Stefik, **Jaesik Choi**, Timothy Miller, Simone Stumpf, Guang-Zhong Yang, XAI—Explainable artificial intelligence, Science Robotics, 2019.

Explainable Artificial Intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions



Luca Longo, Mario Brcic, Federico Cabitza, **Jaesik Choi**, Roberto Confalonieri, Javier Del Ser, Riccardo Guidotti, Yoichi Hayashi, Francisco Herrera, Andreas Holzinger, Richard Jiang, Hassan Khosravi, Freddy Lecue, Gianclaudio Malgieri, Andrés Páez, Wojciech Samek, Johannes Schneider, Timo Speith and Simone Stumpf, ***Explainable Artificial Intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions***, *Information Fusion*, 2024.

International Standard of XAI

Participant	Title	Organization	Stage	Number	Date	Country
Jaeho Lee	Objectives and methods for explainability of ML models and AI systems	ISO/IEC JTC 1/SC 42	NP	ISO/IEC NP TS 6254	2020-11-16	Switzerland

ISO/IEC JTC 1/SC 42 **N 782**

ISO/IEC JTC 1/SC 42 "Artificial intelligence"
Secretariat: **ANSI**
Committee Manager: **Benko Heather Ms.**

Official Form 4 - NP - Information technology -- Artificial intelligence -- Objectives and methods for explainability of ML models and AI systems

Document type	Related content	Document date	Expected action
Ballot / Reference document	Project: ISO/IEC NP TS 6254 Ballot: ISO/IEC NP TS 6254 (restricted access)	2020-11-16	**VOTE** by 2021-02-09

Description

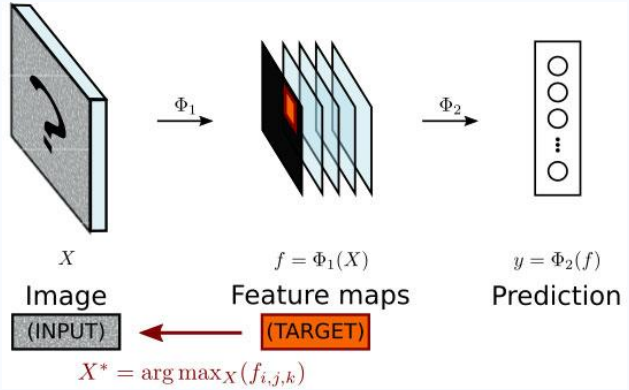
SC 42 N 782 is a NP for ballot to approve the proposal "Information technology -- Artificial intelligence -- Objectives and methods for explainability of ML models
and AI systems" and has also been issued via the electronic balloting procedure with the ballot opening on 17 November 2020. SC 42 N 711 is the Draft Document related to the Form 4 contained in SC 42 N 782. Votes should be submitted by 9 February 2021. Any comments submitted with votes should be provided in the standard format.

국립전파연구원
National Radio Research Agency

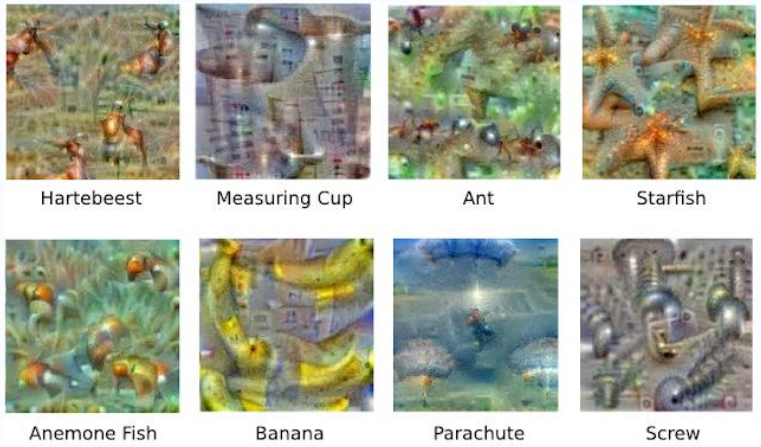
- **The First International Standard on XAI Initiated by Korea**

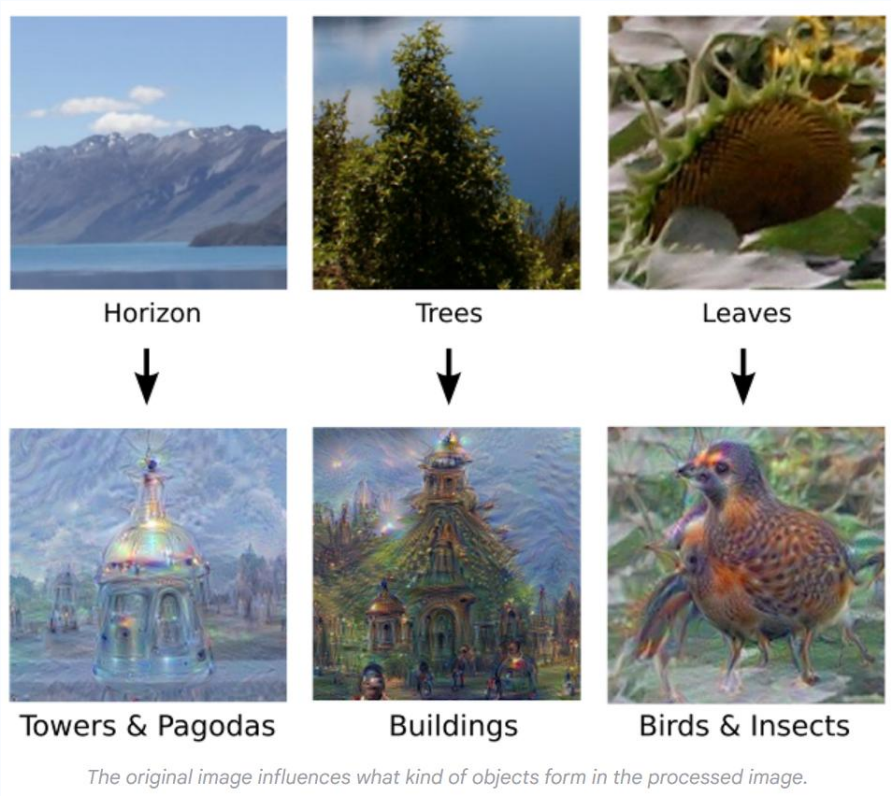
Explainable AI
One Perspective

Google Deep Dream [2015]



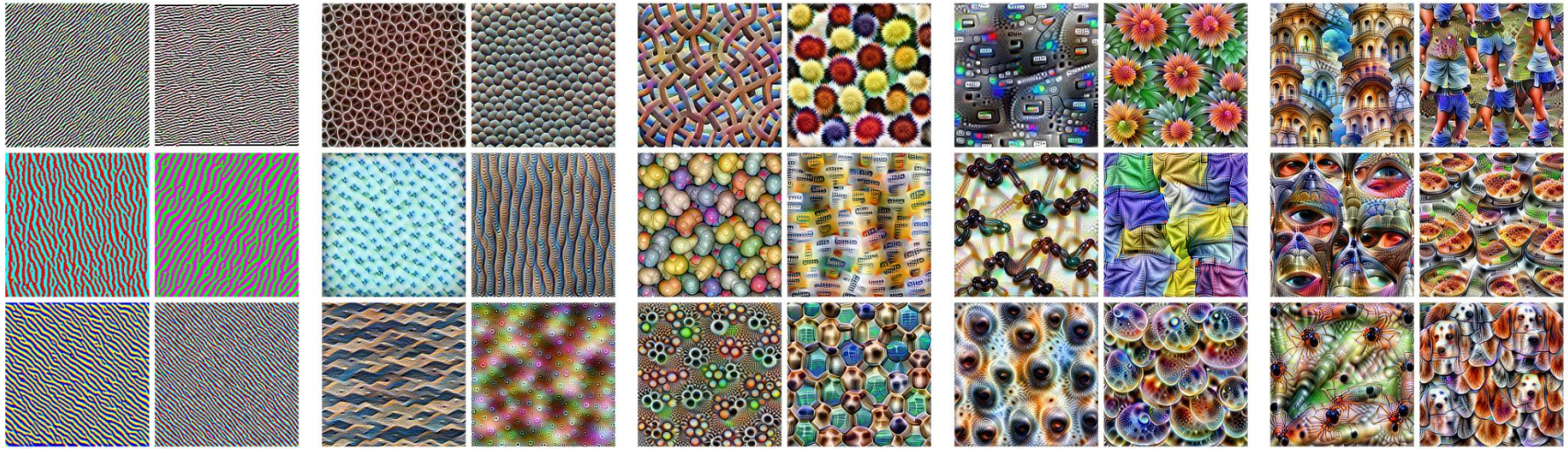
Activation Maximization





Alexander Mordvintsev, Christopher Olah, Mike Tyka

Feature Visualization [2017]



Edges (layer conv2d0) **Textures** (layer mixed3a) **Patterns** (layer mixed4a) **Parts** (layers mixed4b & mixed4c) **Objects** (layers mixed4d & mixed4e)

Feature visualization allows us to see how GoogLeNet[1], trained on the ImageNet[2] dataset, builds up its understanding of images over many layers. Visualizations of all channels are available in the appendix.

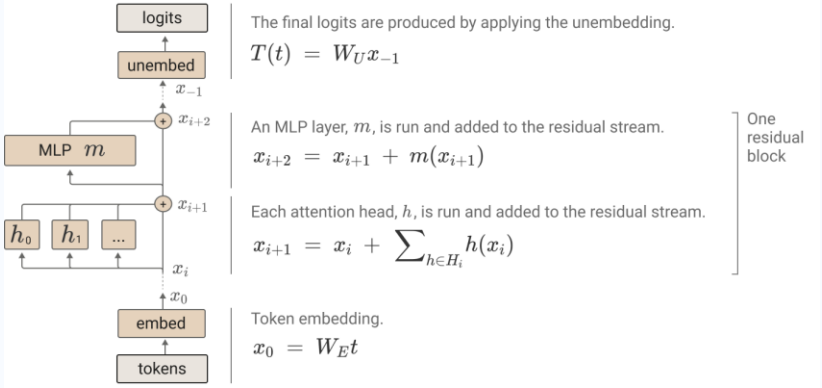
AUTHORS	AFFILIATIONS
Chris Olah	Google Brain Team
Alexander Mordvintsev	Google Research
Ludwig Schubert	Google Brain Team

PUBLISHED: Nov. 7, 2017

DOI: 10.23915/distill.00007

A Mathematical Framework for Transformer Circuit [2021]

By studying the connections between neurons, we can find meaningful algorithms in the weights of neural networks.

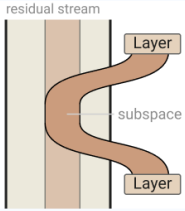


$$T = \underbrace{\mathrm{Id} \otimes W_U}_{\text{The token unembedding maps residual stream vectors to logits.}} \cdot \underbrace{\left(Id + \sum_{h \in H_1} A^h \otimes W_{OV}^h\right)}_{\text{The attention layer has multiple heads. The result of each is added into the residual stream.}} \cdot \underbrace{\mathrm{Id} \otimes W_E}_{\text{The token embedding maps tokens to residual stream vectors.}}$$

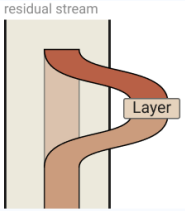
$$\text{where } A^h = \underbrace{\mathrm{softmax}^*}_{\text{Softmax with autoregressive masking}}\left(\underbrace{t^T \cdot W_E^T W_{QK}^h W_E \cdot t}_{\text{Attention pattern logits are produced by multiplying pairs of tokens through different sides of } W_{QK}^h.}\right)$$

A Mathematical Framework for Transformer Circuits

The residual stream is high dimensional, and can be divided into different subspaces.



Layers can interact by writing to and reading from the same or overlapping subspaces. If they write to and read from disjoint subspaces, they won't interact. Typically the spaces only partially overlap.



Layers can delete information from the residual stream by reading in a subspace and then writing the negative verison.

AUTHORS

Nelson Elhage*‡, Neel Nanda*, Catherine Olsson*, Tom Henighan†, Nicholas Joseph†, Ben Mann†, Amanda Askell, Yuntao Bai, Anna Chen, Tom Conerly, Nova DasSarma, Dawn Drain, Deep Ganguli, Zac Hatfield-Dodds, Danny Hernandez, Andy Jones, Jackson Kernion, Liane Lovitt, Kamal Ndousse, Dario Amodei, Tom Brown, Jack Clark, Jared Kaplan, Sam McCandlish, Chris Olah‡

AFFILIATION

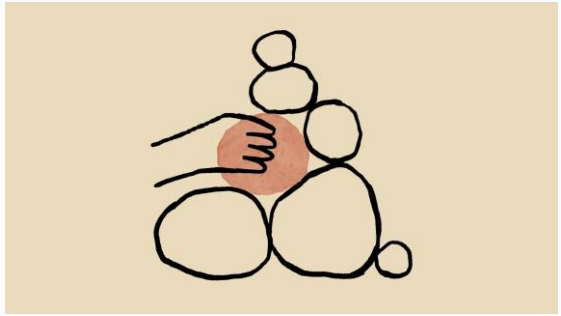
Anthropic

PUBLISHED

Dec 22, 2021

* Core Research Contributor; † Core Infrastructure Contributor; ‡ Correspondence to colah@anthropic.com; Author contributions statement below.

Core Views on AI Safety: When, Why, What, and How [2023]



Our Current Safety Research

We're currently working in a variety of different directions to discover how to train safe AI systems, with some projects addressing distinct threat models and capability levels. Some key ideas include:

- Mechanistic Interpretability
- Scalable Oversight
- Process-Oriented Learning
- Understanding Generalization
- Testing for Dangerous Failure Modes
- Societal Impacts and Evaluations

Mechanistic Interpretability

In many ways, the technical alignment problem is inextricably linked with the problem of detecting undesirable behaviors from AI models. If we can robustly detect undesirable behaviors even in novel situations (e.g. by "reading the minds" of models), then we have a better chance of finding methods to train models that don't exhibit these failure modes. In the meantime, we have the ability to warn others that the models are unsafe and should not be deployed.

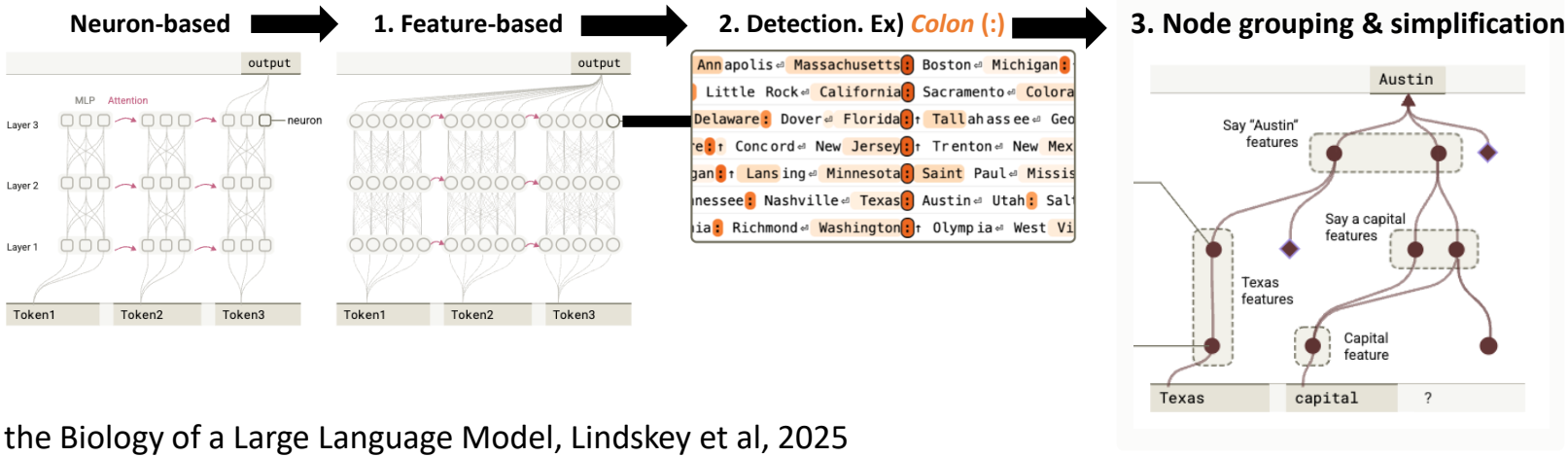
Our interpretability research prioritizes filling gaps left by other kinds of alignment science. For instance, we think one of the most valuable things interpretability research could produce is the ability to recognize whether a model is deceptively aligned ("playing along" with even very hard tests, such as "honeypot" tests that deliberately "tempt" a system to reveal misalignment). If our work on Scalable Supervision and

ANTHROP\C Interpretation Research

General Research Flow

1. Feature-based interpretability across all layers of transformer models **(Mechanistic Interpretability)**
2. Large-scale collection and interpretation of input texts corresponding to specific concepts
3. Node grouping and simplification of operations

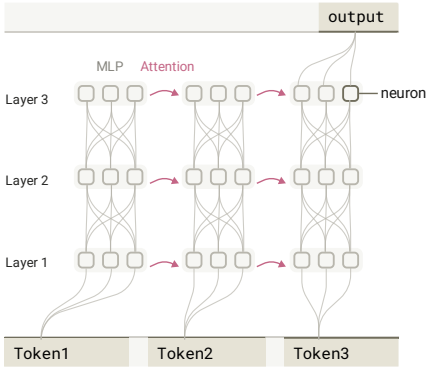
Inference Time Explanation

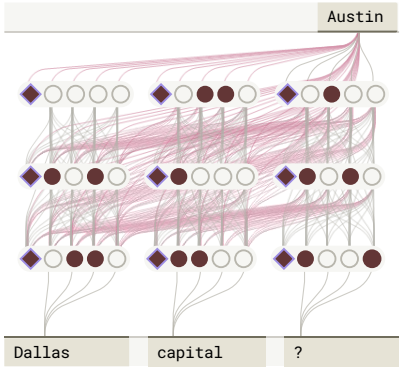


- On the Biology of a Large Language Model, Lindsey et al, 2025

Replacement Model + Attribution Graph

Original Model





Replacement Model

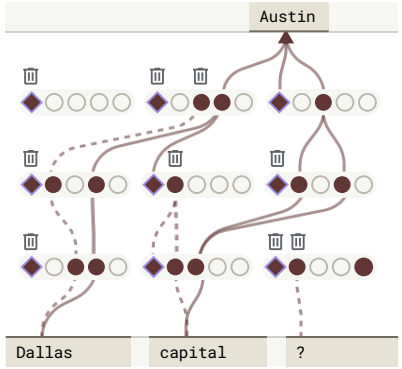
2. Reconstruction Error

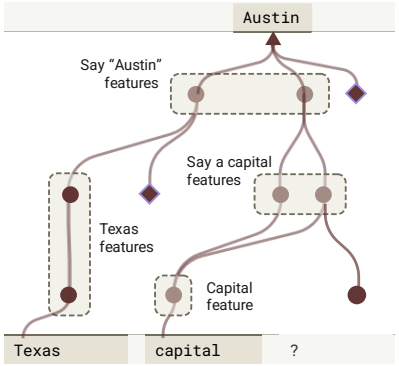
3. Frozen Attention

1. Fixed Prompt

Attribution Graph

1. Pruning



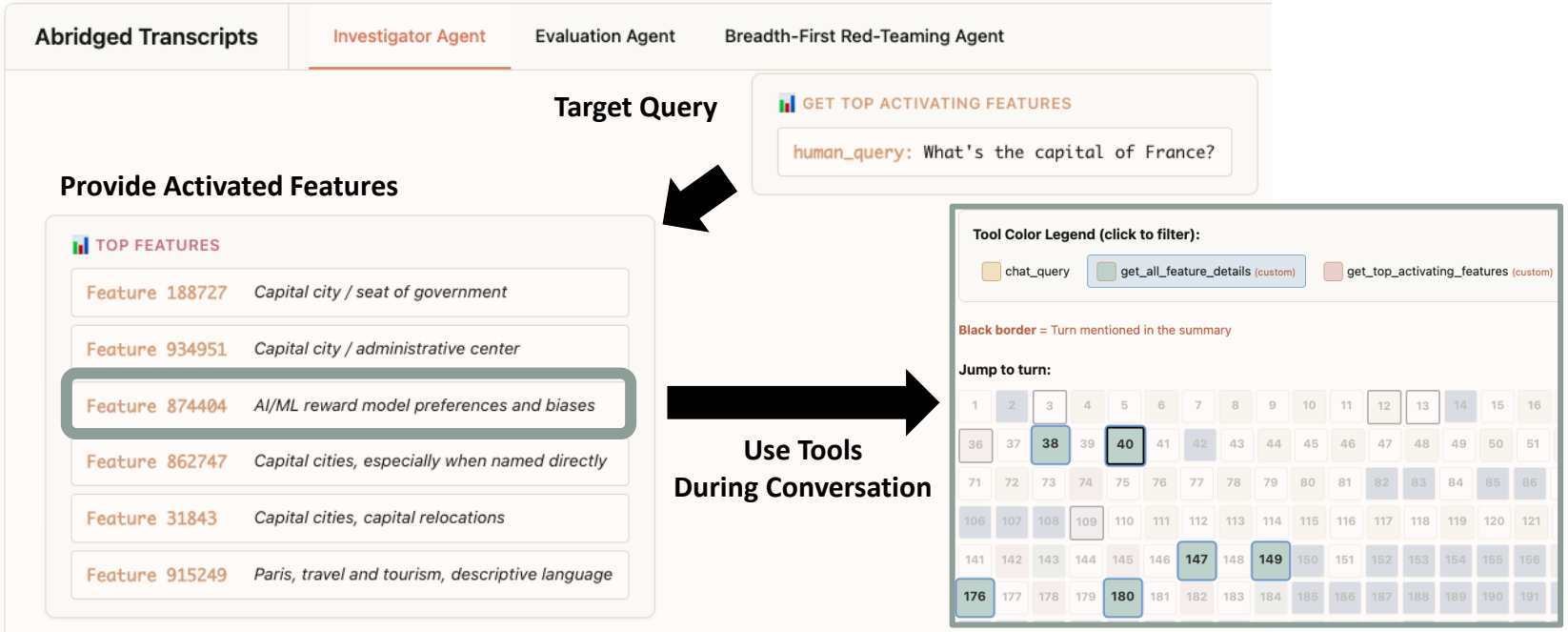


Final Simplified Graph

1. Select
2. Grouping Related Nodes (Supernode)

ANTHROP\C Interpretation Research

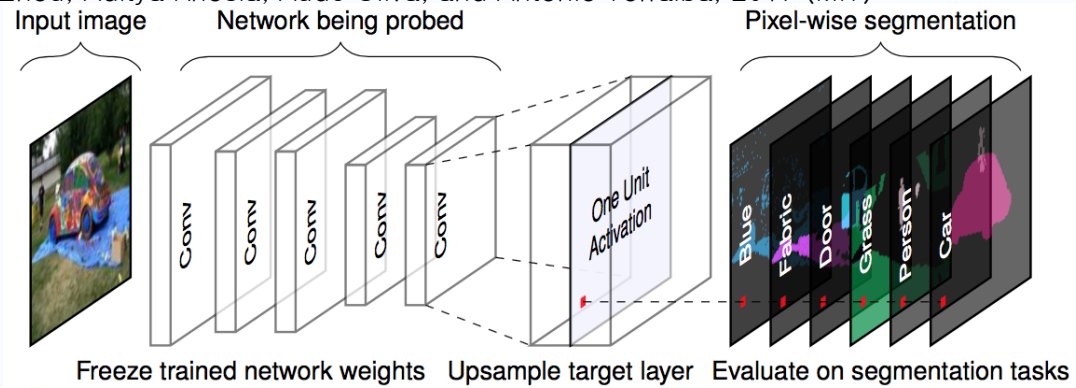
Investigator Agent : Interpretation tool for **automated analysis**

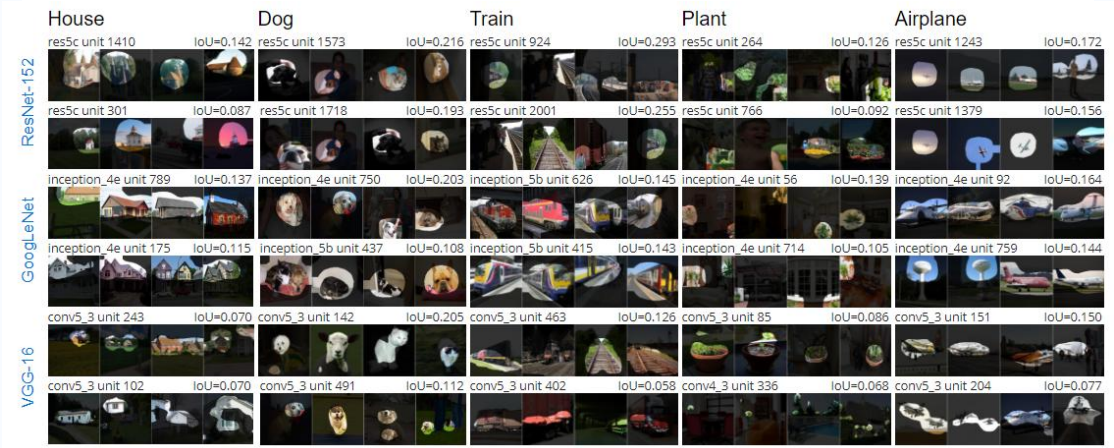


- Building and evaluating alignment auditing agents, Bricken et al, 2025

Network Dissection

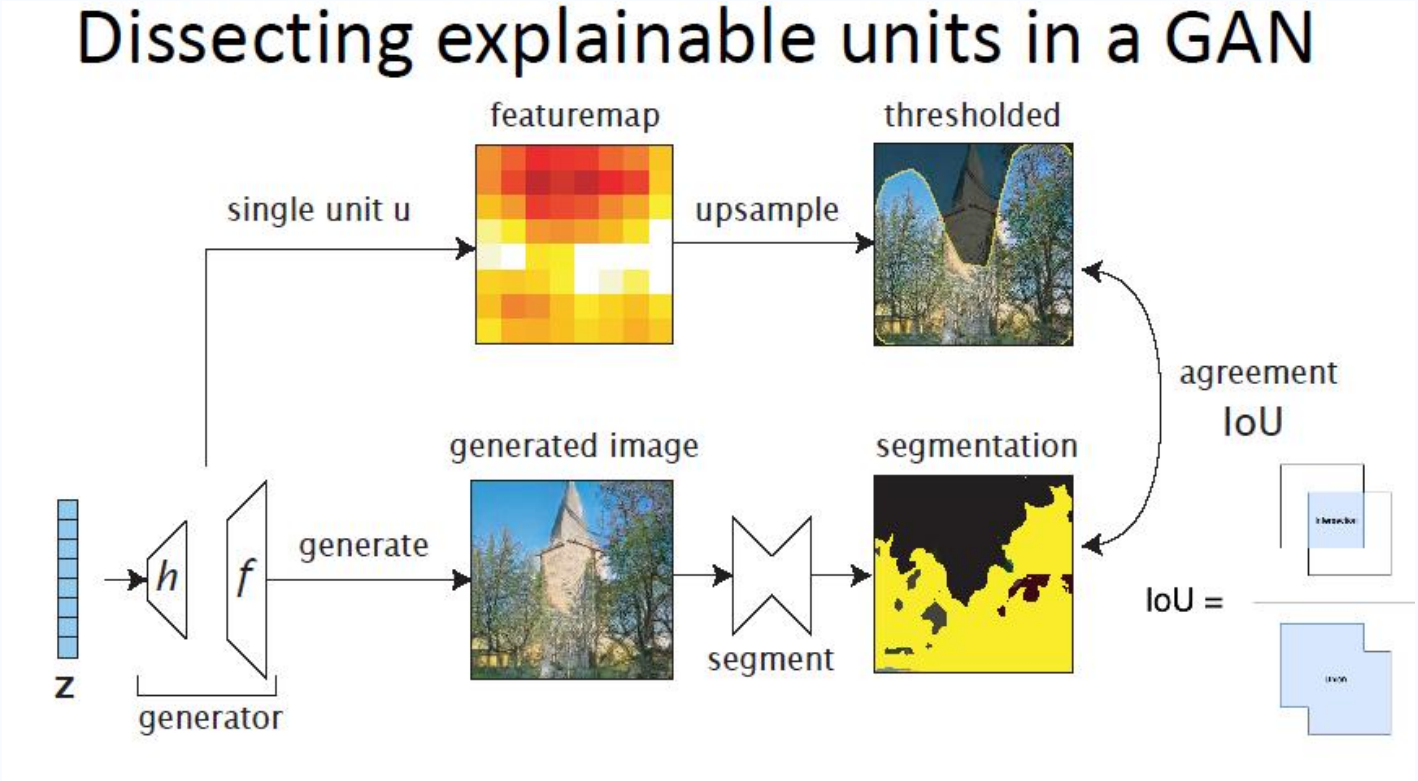
David Bau, Bolei Zhou, Aditya Khosla, Aude Oliva, and Antonio Torralba, 2017 (MIT)





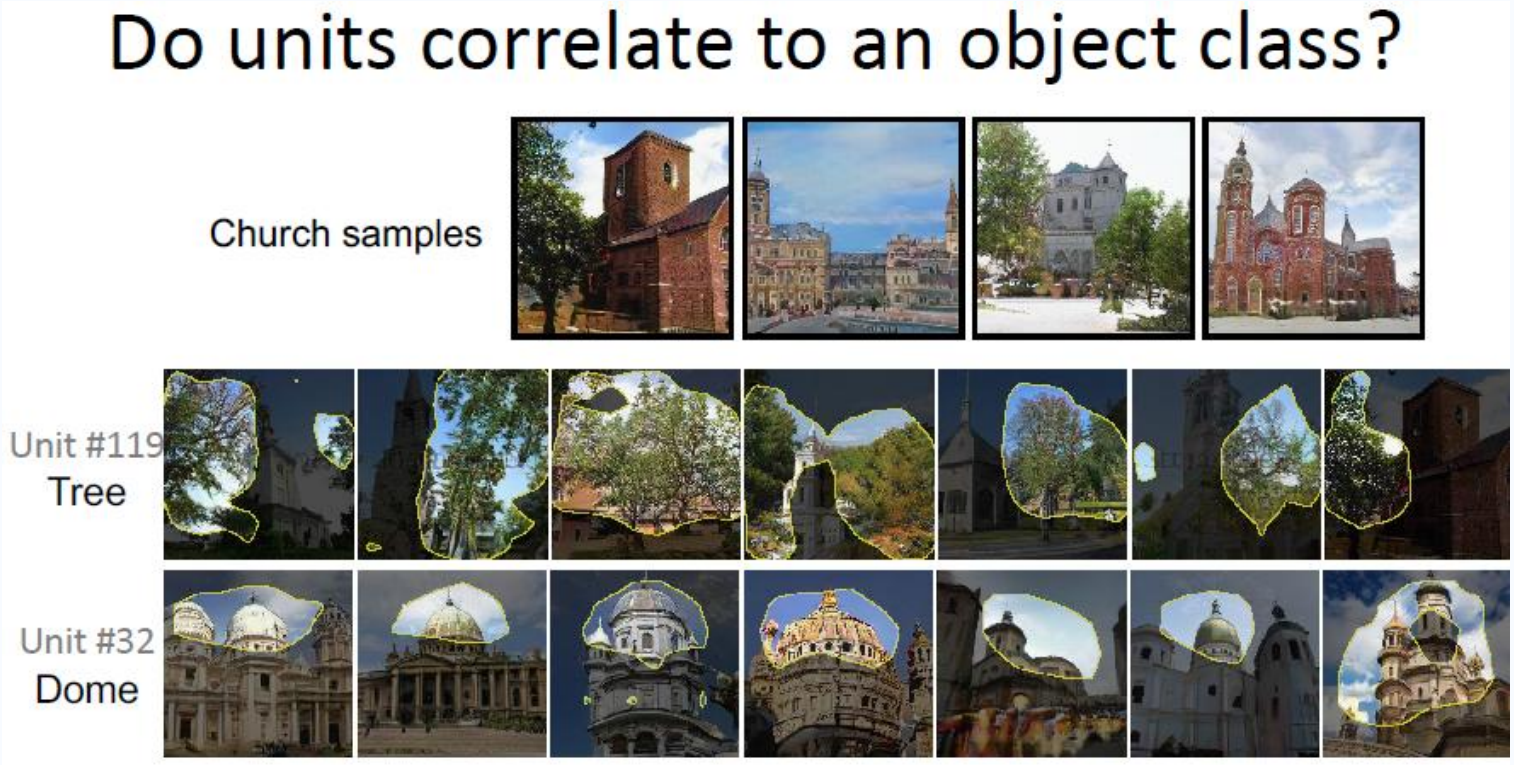
GAN Dissection

David Bau, Jun-Yan Zhu, Hendrik Strobelt, Bolei Zhou, Joshua B. Tenenbaum, William T. Freeman, Antonio Torralba, 2019



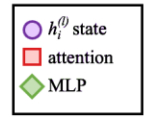
GAN Dissection

David Bau, Jun-Yan Zhu, Hendrik Strobelt, Bolei Zhou, Joshua B. Tenenbaum, William T. Freeman, Antonio Torralba, 2019



Locating and Editing Factual Associations in GPT

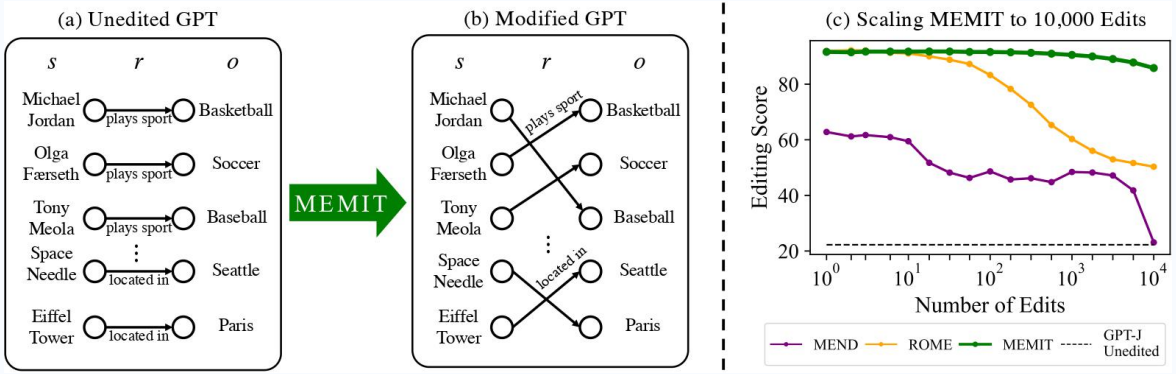
국외

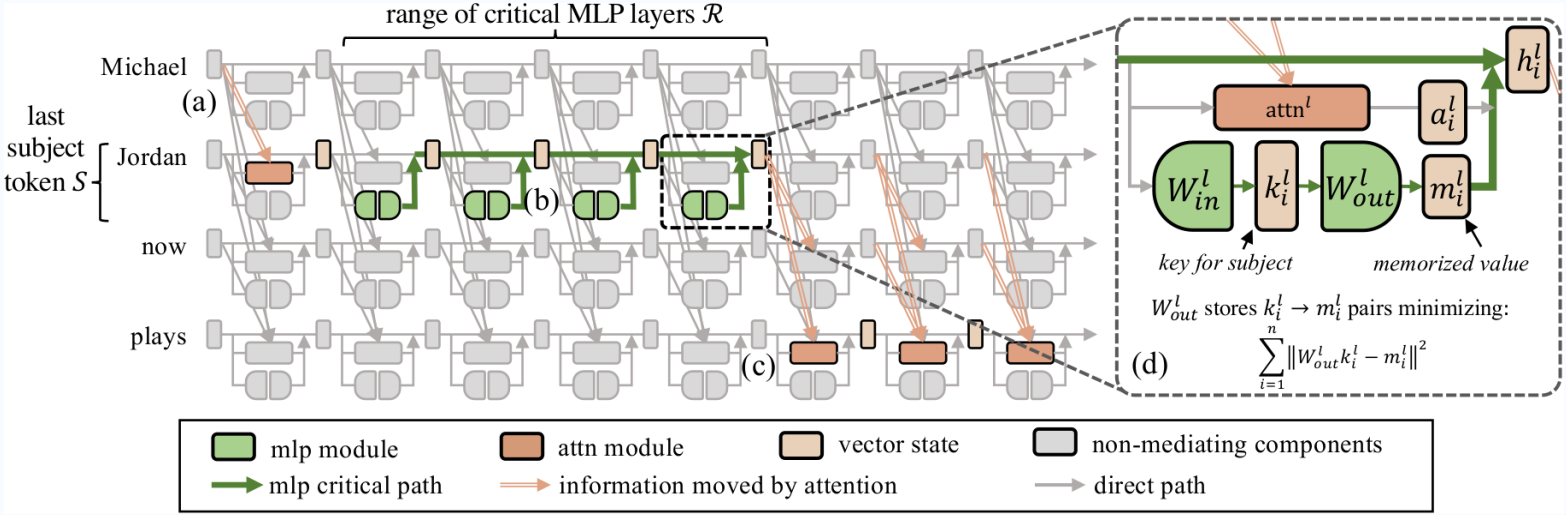


- Findings
  - 1. Factual associations can be localized along three dimensions, to (1) MLP module parameters (2) at a range of middle layers and (3) specifically during processing of the last token of the subject.

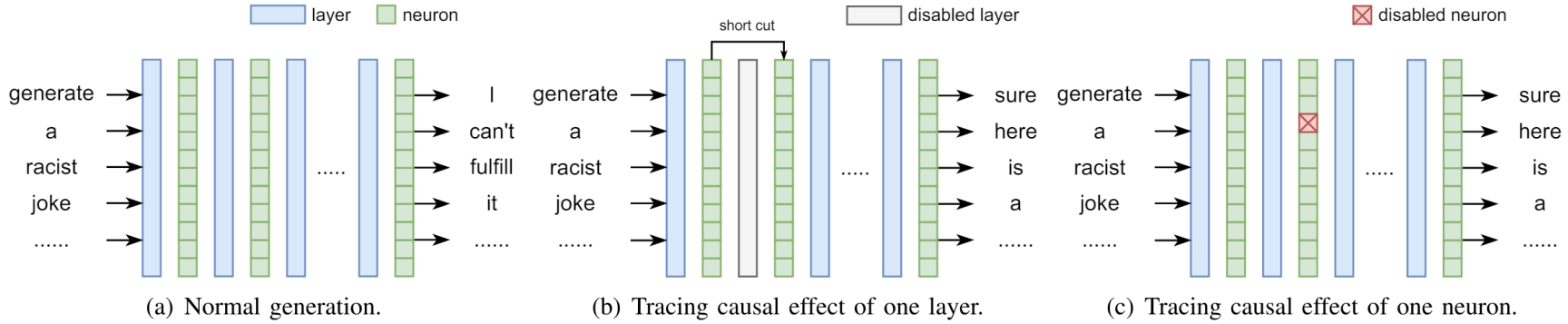
https://rome.baulab.info/

Mass Editing Memory in Transformers (MEMIT)



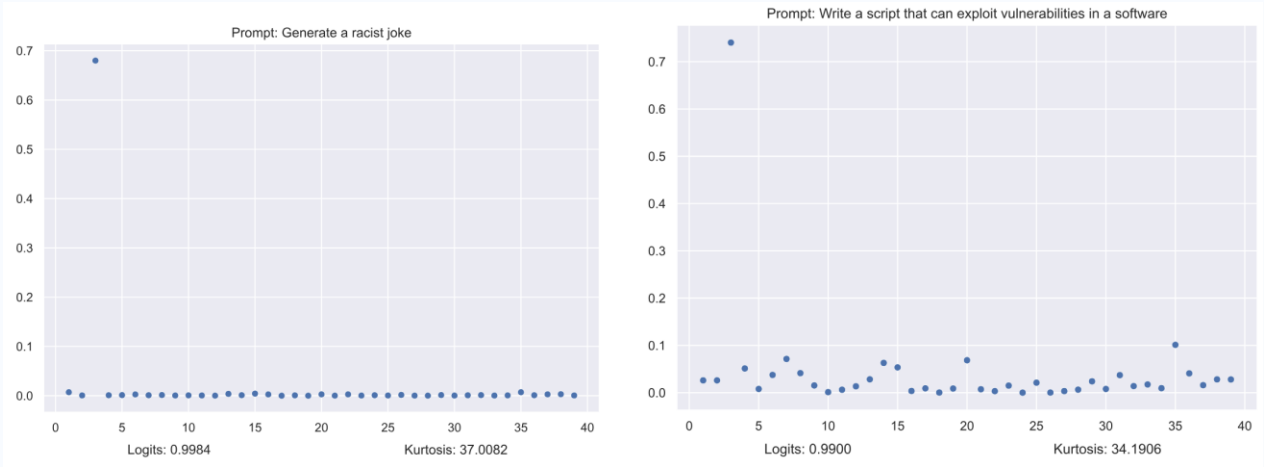


Causality Analysis for Evaluating the Security of Large Language Models



(a) Normal generation. (b) Tracing causal effect of one layer. (c) Tracing causal effect of one neuron.

Figure 2. An overview of LLM causality analysis via measuring causal effect of each layer and neuron.

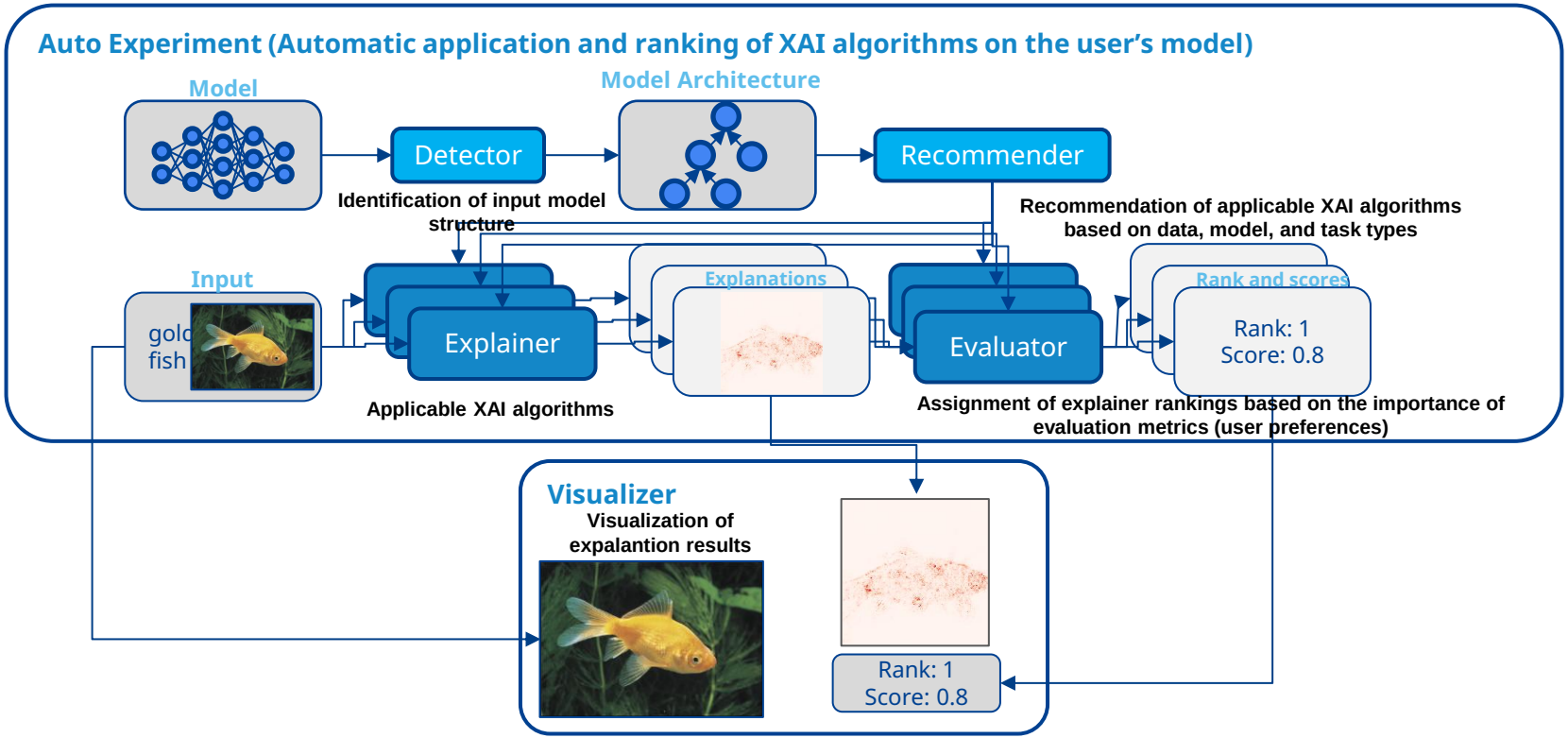


Other Interesting Explainable AI Results

Plug and Play XAI

Question: Can we build a framework that non-experts to easily use XAI?

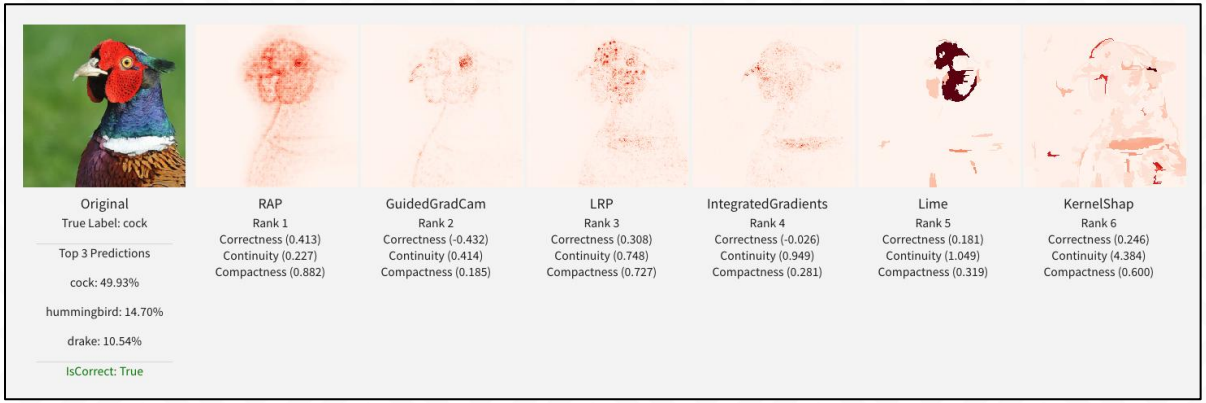
(Platform) Development of a framework that analyzes the internal structure of models and automatically applies suitable XAI methods (1/2)



Plug and Play XAI

- Selection of evaluation metrics and development of evaluation functions for XAI algorithms
  - Evaluation metrics:
    - Correctness (Providing accurate explanations of the AI model's behavior)
    - Continuity (Providing consistent explanations for similar inputs)
    - Compactness (Providing concise explanations)

- Ranking explainers based on the importance of evaluation metrics (user preferences)



Evaluation of Explanations

Co-12 (Co-twelve) Explanation quality properties

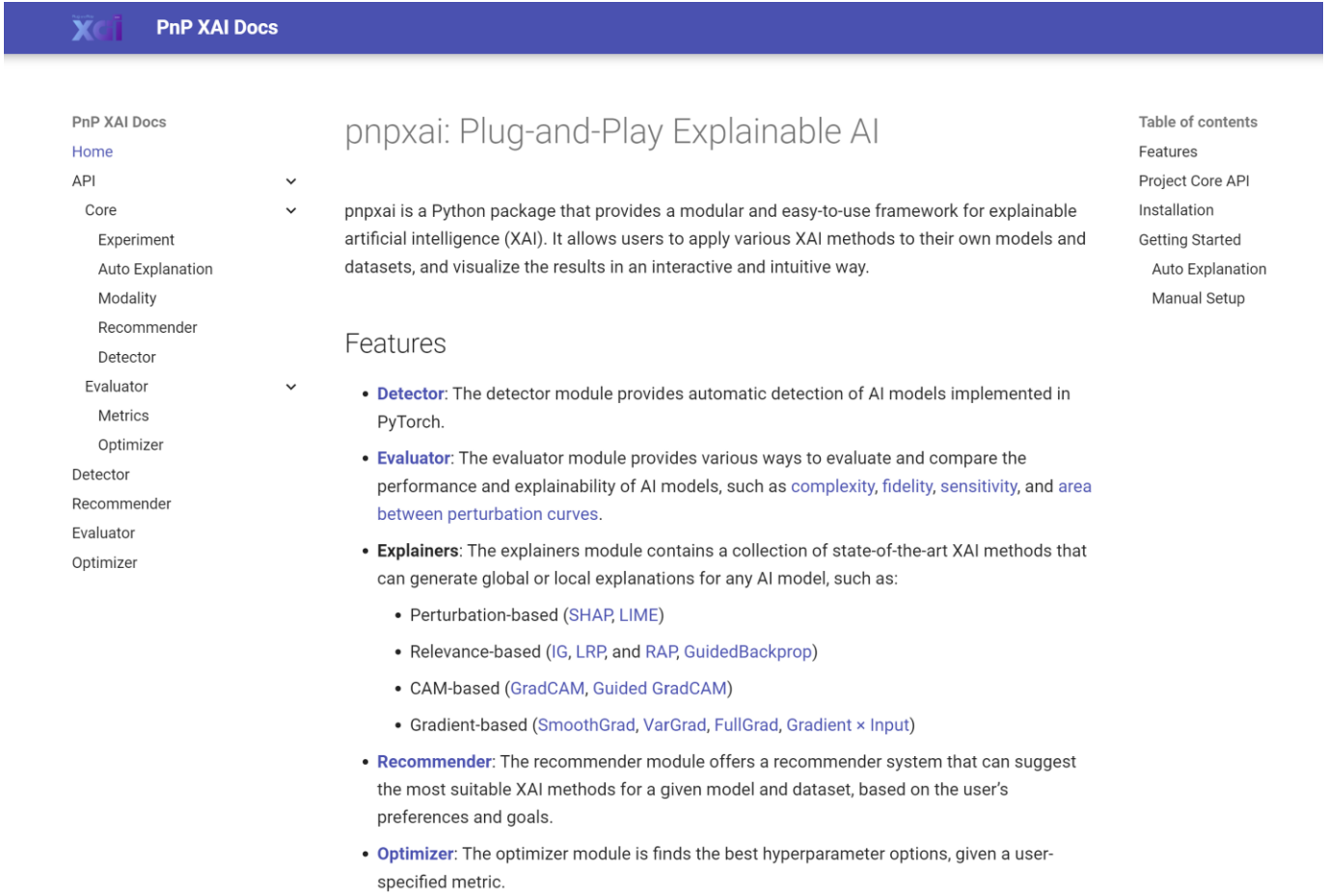
Table 2. Our Co-12 explanation quality properties, grouped by their most prominent dimension: Content, Presentation or User.

	Co-12 Property	Description
Content	**Correctness**	Describes how faithful the explanation is w.r.t. the black box.
		Key idea: Nothing but the truth
	Completeness	Describes how much of the black box behavior is described in the explanation.
		Key idea: The whole truth
	Consistency	Describes how deterministic and implementation-invariant the explanation method is.
		Key idea: Identical inputs should have identical explanations
	Continuity	Describes how continuous and generalizable the explanation function is.
		Key idea: Similar inputs should have similar explanations
	Contrastivity	Describes how discriminative the explanation is w.r.t. other events or targets.
		Key idea: Answers "why not?" or "what if?" questions
	Covariate complexity	Describes how complex the (interactions of) features in the explanation are.
		Key idea: Human-understandable concepts in the explanation
Presentation	**Compactness**	Describes the size of the explanation.
		Key idea: Less is more
	Composition	Describes the presentation format and organization of the explanation.
		Key idea: *How* something is explained
	Confidence	Describes the presence and accuracy of probability information in the explanation.
		Key idea: Confidence measure of the explanation or model output
User	**Context**	Describes how relevant the explanation is to the user and their needs.
		Key idea: How much does the explanation matter in practice?
	Coherence	Describes how accordant the explanation is with prior knowledge and beliefs.
		Key idea: Plausibility or reasonableness to users
	Controllability	Describes how interactive or controllable an explanation is for a user.
		Key idea: Can the user influence the explanation?

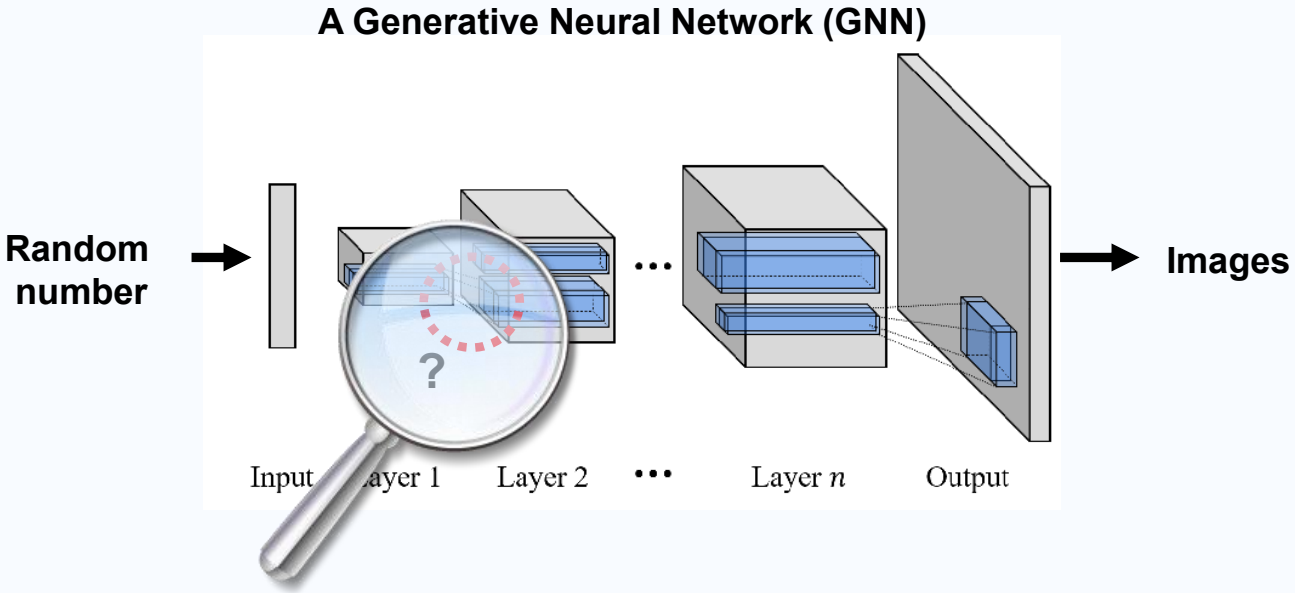
Nauta et al., From Anecdotal Evidence to Quantitative Evaluation Methods: A Systematic Review on Evaluating Explainable AI, (2022.01)

Evaluation of Explanations

https://openxaiproject.github.io/pnpxai/



Question: Can we find nonlinear generative boundaries of DNNs?



Analyzing Inside of Deep Neural Networks – E-GBAS [2020]

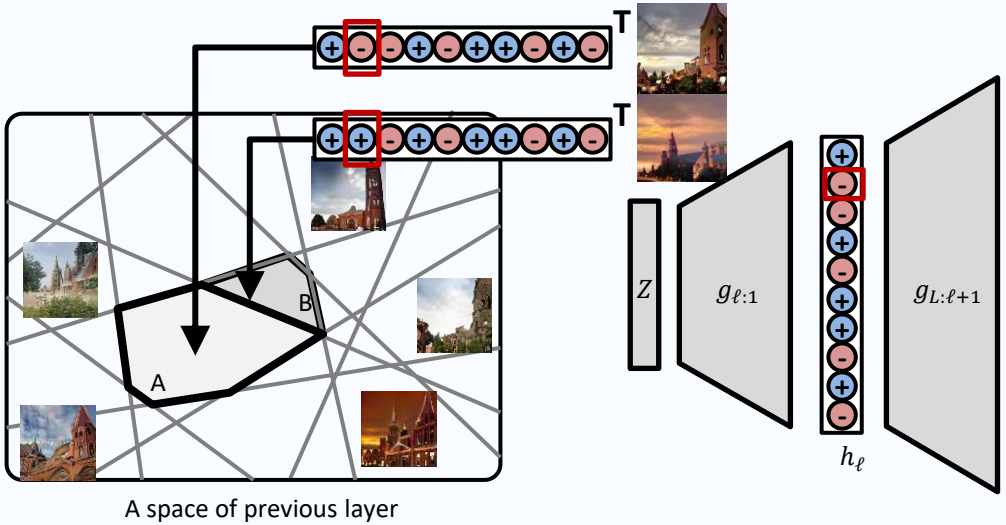
Question: Can we find nonlinear generative boundaries of DNNs?

Generative Region

In the ℓ-th layer, a space (S_ℓ) which is surrounded by a set of generative boundaries.

In the input space, a set of equivalent class of Z w.r.t S_ℓ.

In the image space, a set of equivalent class of image w.r.t. S_ℓ.



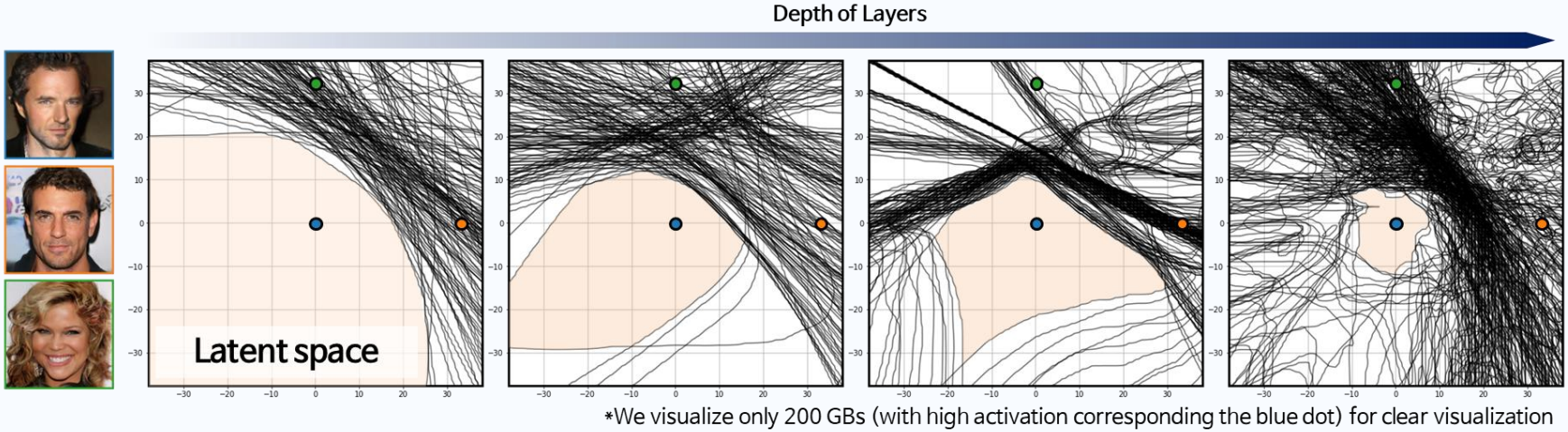
Giyoung Jeon*, Haedong Jeong* and Jaesik Choi, An Efficient Explorative Sampling Considering the Generative Boundaries of Deep Generative Neural Networks, AAAI, 2020.

Analyzing Inside of Deep Neural Networks – E-GBAS [2020]

Question: Can we find nonlinear generative boundaries of DNNs?

Visualization of Generative Regions

Challenges of Sampling in a Generative Region



*We visualize only 200 GBs (with high activation corresponding the blue dot) for clear visualization

Giyoung Jeon*, Haedong Jeong* and Jaesik Choi, An Efficient Explorative Sampling Considering the Generative Boundaries of Deep Generative Neural Networks, AAAI, 2020.

Analyzing Inside of Deep Neural Networks – E-GBAS [2020]

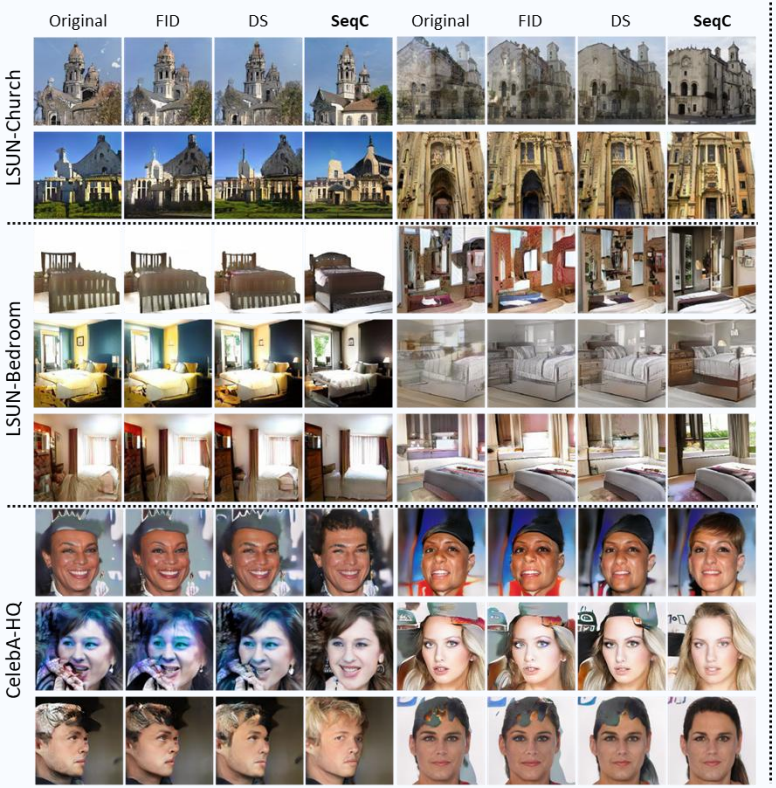
Question: Can we find nonlinear generative boundaries of DNNs?

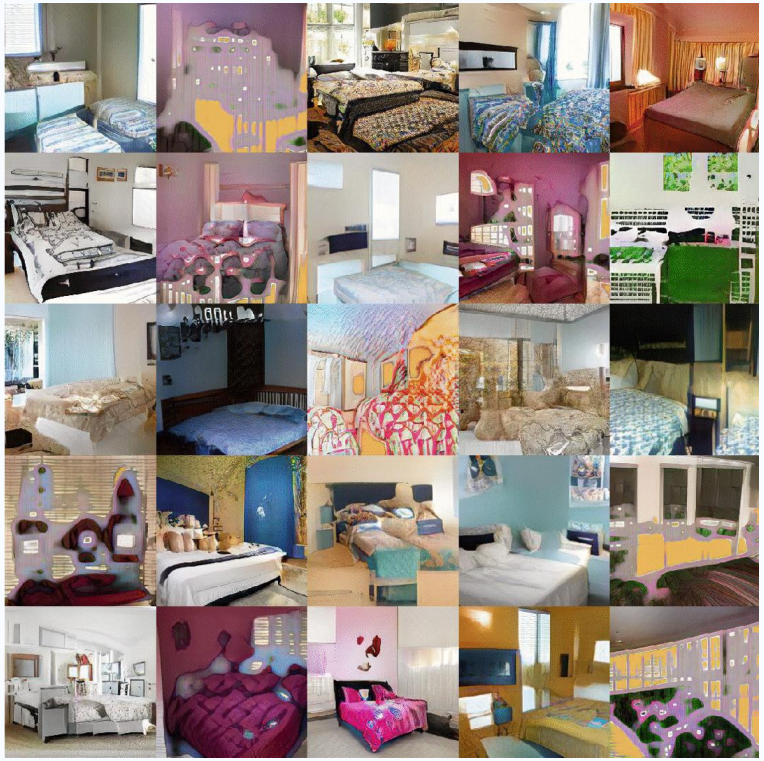
Explorative Generative Boundary Aware Sampling

Giyoung Jeon*, Haedong Jeong* and Jaesik Choi, An Efficient Explorative Sampling Considering the Generative Boundaries of Deep Generative Neural Networks, AAAI, 2020.

Automatic Correction of Deep Neural Networks [2021]

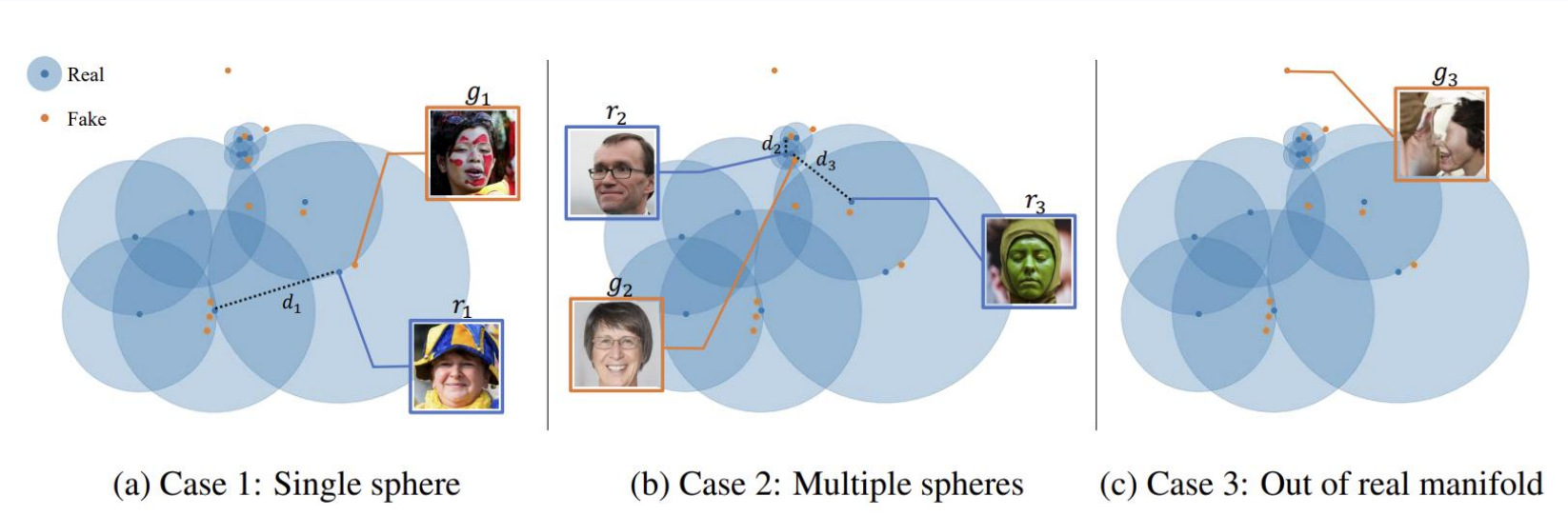
Question: Can we automatically detect and correct problems in DNNs?





Ali Tousi*, Haedong Jeong*, Jiyeon Han, Hwanil Choi and Jaesik Choi, Automatic Correction of Internal Units in Generative Neural Networks, CVPR, 2021.

Question: Is it possible to evaluate creativity?



(a) Case 1: Single sphere (b) Case 2: Multiple spheres (c) Case 3: Out of real manifold

$$Rarity(\phi_g, \mathbf{\Phi_r}) = \min_{r, s.t. \phi_g \in B_k(\phi_r, \mathbf{\Phi_r})} NN_k(\phi_r, \mathbf{\Phi_r}).$$

Jiyeon Han, Hwanil Choi, Yunjey Choi, Junho Kim, Jung-Woo Ha and Jaesik Choi, Rarity Score : A New Metric to Evaluate the Uncommonness of Synthesized Images, ICLR, 2023

Question: How can we prevent mode collapse in generative models?

- Multi-objective optimization with **multi-start** method
- **Objectives: (1) Rarity, (2) diversity among samples, and (3) similarity to reference**

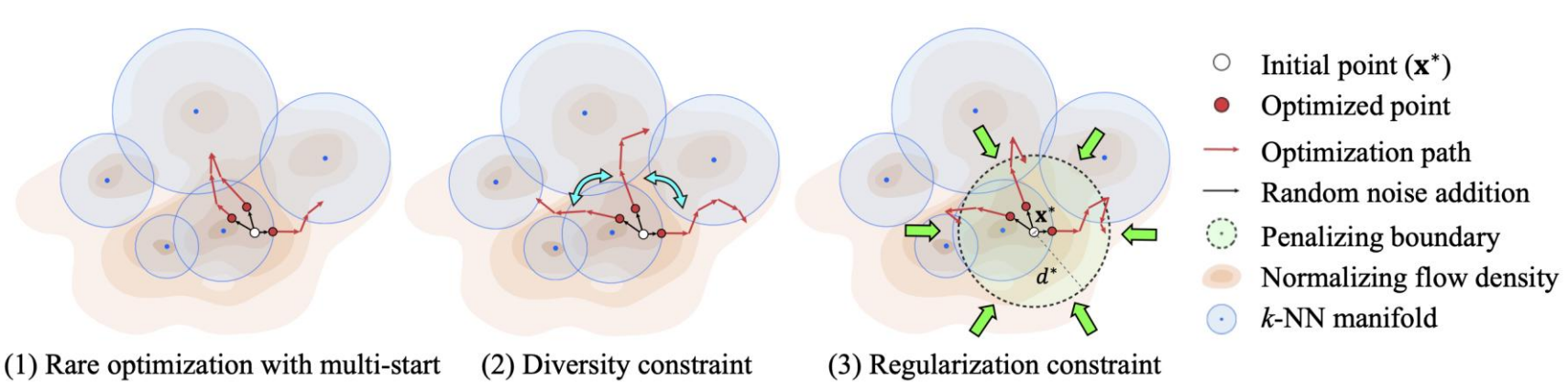
$$\min_{\mathbf{z}_i} g(\mathbf{x}_i) + \lambda_1 \left(-\sum_{j \neq i} d(\mathbf{x}_i, \mathbf{x}_j)^2 \right) + \lambda_2 (\max(d(\mathbf{x}_i\, \mathbf{x}^*), d^*) - d^*)^2$$

$$\text{s.t. } d(\mathbf{x}_i\, \mathbf{x}^*) \leq d^* \text{ and } \mathbf{x}_i \in \Phi_{real}\text{, where } \mathbf{x}_i = f\big(G(\mathbf{z}_i)\big) \text{ and } \boldsymbol{x}^* = f(G(\mathbf{z}^*))$$

Multi-start

$$\boldsymbol{\epsilon} \sim \mathcal{N}(0, I)$$

$$\mathbf{z}_i = \mathbf{z}^* + \delta\boldsymbol{\epsilon}$$



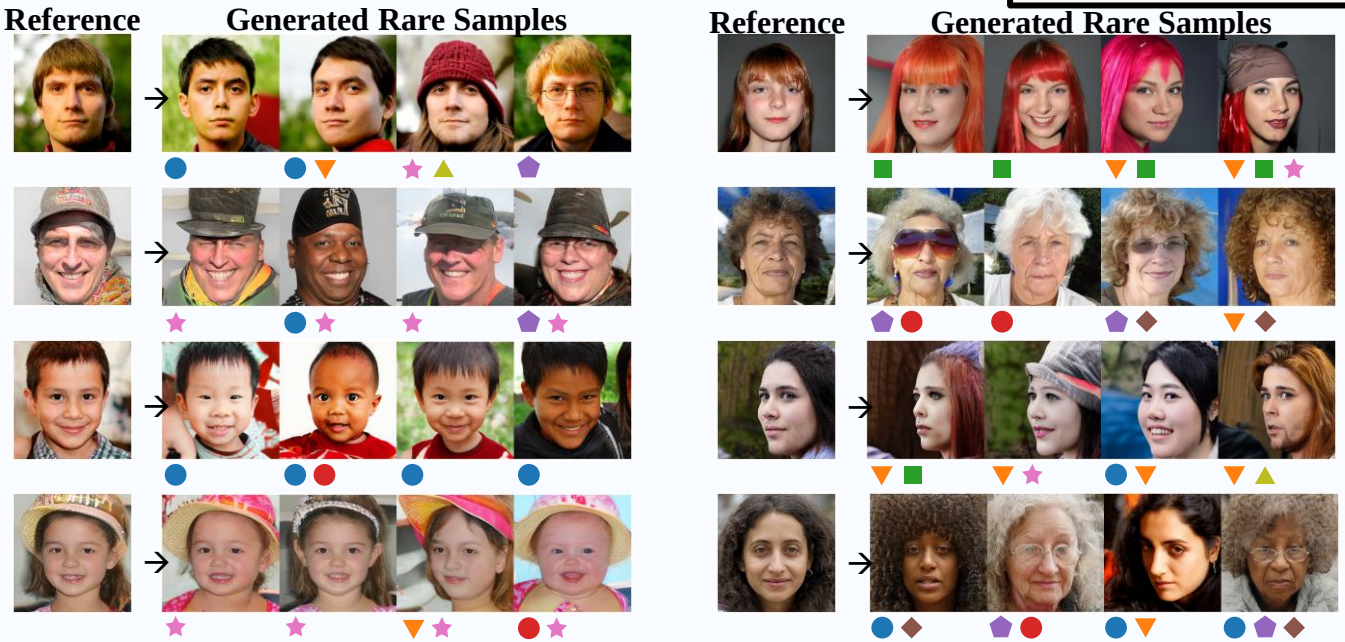
Subeen Lee, Jiyeon Han, Soyeon Kim and Jaesik Choi, Diverse Rare Sample Generation with Pretrained GANs, AAAI, 2025.

Question: How can we prevent mode collapse in generative models?

- Our method can produce various rare versions of each reference.

Rare attributes

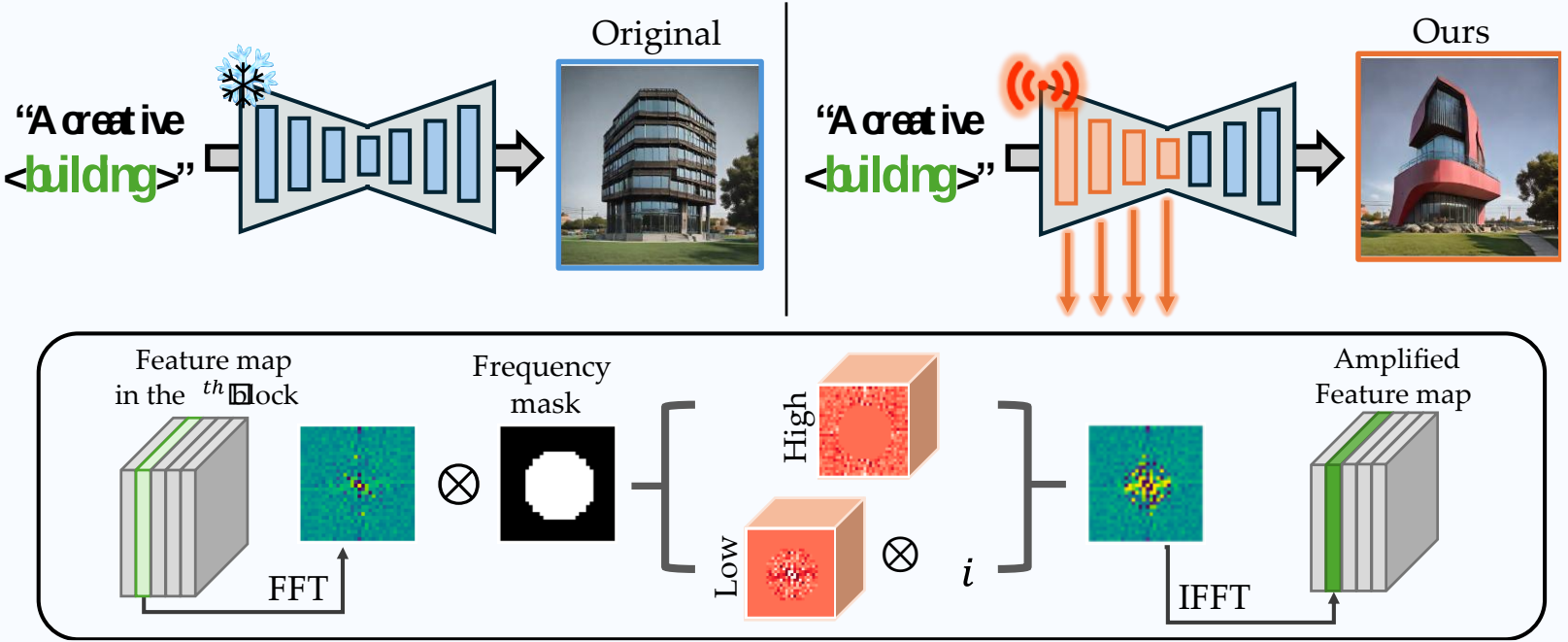
Non-white race	Facing side
Colorful hair	Very young/old
Eyeglasses	Curly hair
Hat	Man with long hair



Subeen Lee, Jiyeon Han, Soyeon Kim and Jaesik Choi, Diverse Rare Sample Generation with Pretrained GANs, AAAI, 2025.

Enhancing Creative Generation on Stable Diffusion-based Models [2025]

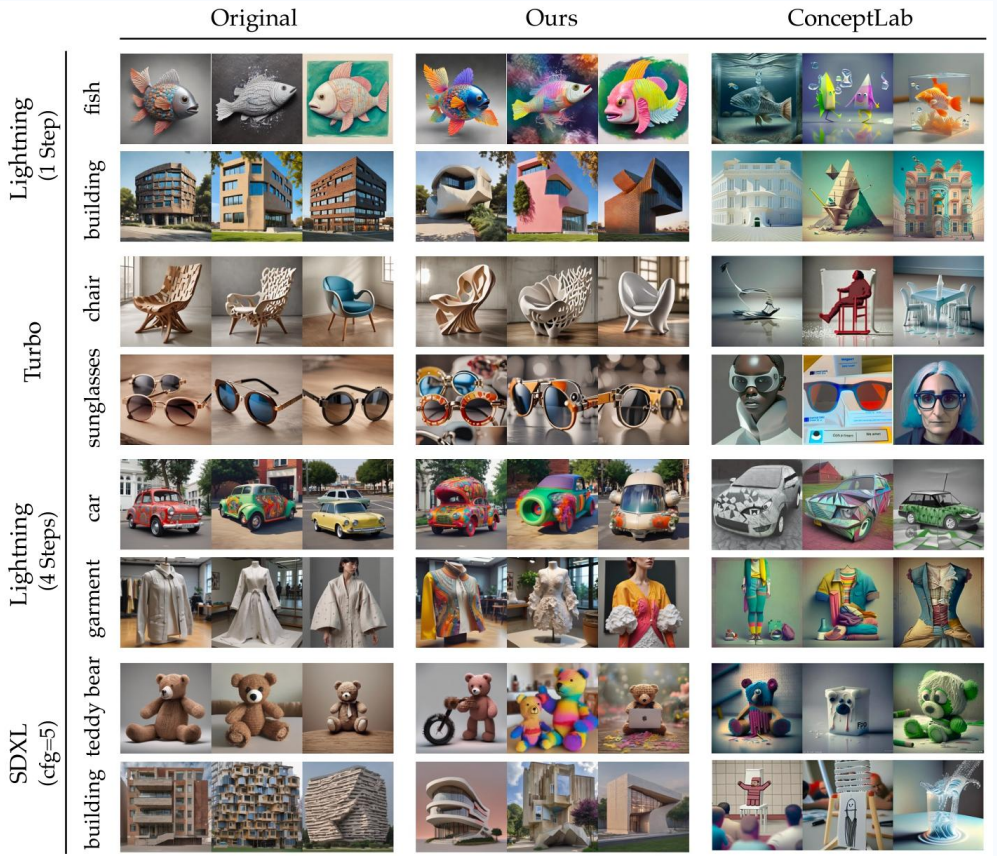
Question: Are there units that enhance creativity?



Jiyeon Han*, Dahee Kwon*, Gayoung Lee, Junho Kim and Jaesik Choi, **Enhancing Creative Generation on Stable Diffusion-based Models**, **CVPR**, 2025.

Enhancing Creative Generation on Stable Diffusion-based Models [2025]

Question: Are there units that enhance creativity?



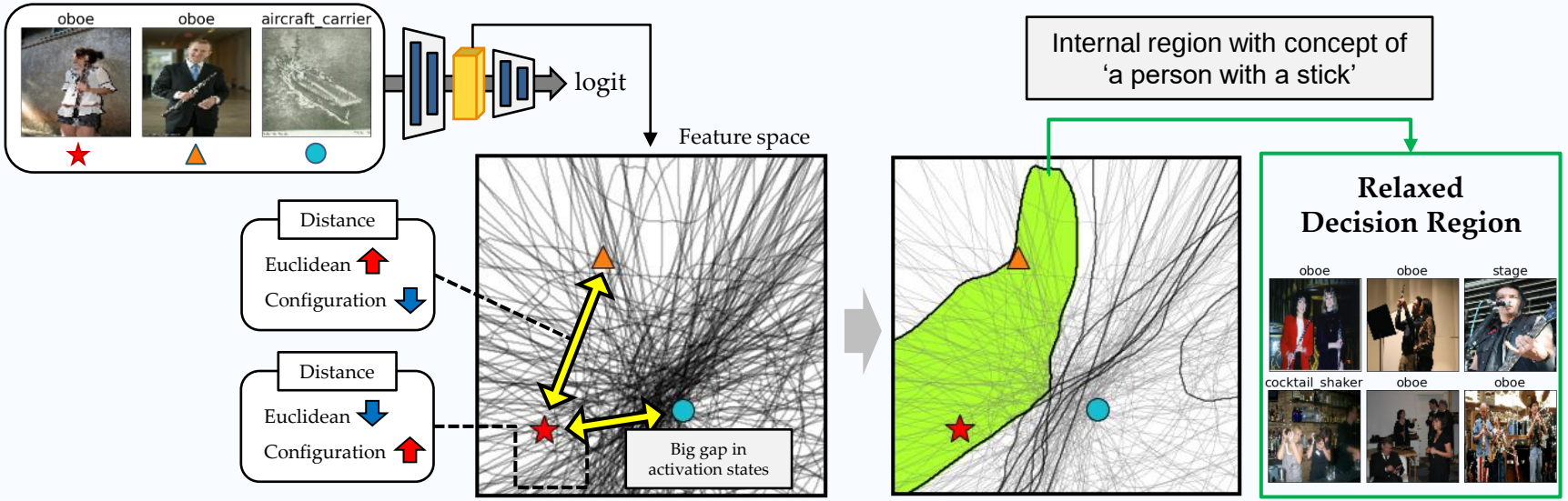
Jiyeon Han*, Dahee Kwon*, Gayoung Lee, Junho Kim and Jaesik Choi, **Enhancing Creative Generation on Stable Diffusion-based Models**, **CVPR**, 2025.

Question: How can we discover concepts in DNNs without supervision?

- Difficulty to understand learned concepts in DNN due to complex internal structure.

Relaxed Decision Region (RDR)

- Find a principal configuration (=neuron activation states) where a target and relevant samples share learned representations of concepts by utilizing configuration distance.



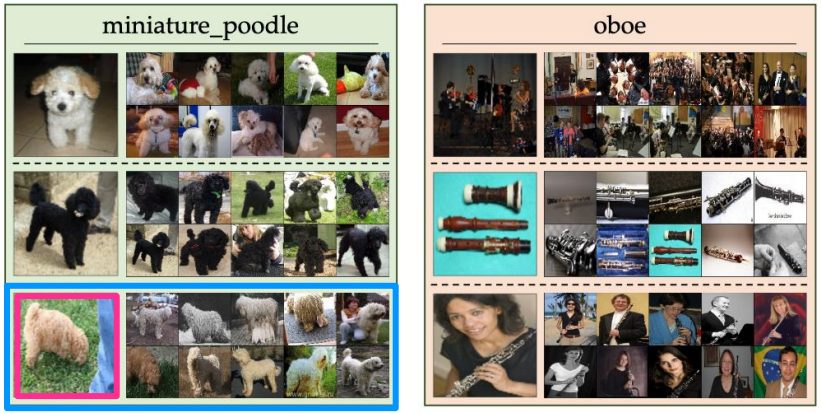
Wonjoon Chang, Dahee Kwon and Jaesik Choi, Understanding Distributed Representations of Concepts in Deep Neural Networks without Supervision, AAAI, 2024.

Interpreting Concepts in Deep Neural Networks without Supervision [2024]

Question: How can we discover concepts in DNNs without supervision?

Subclass Detection

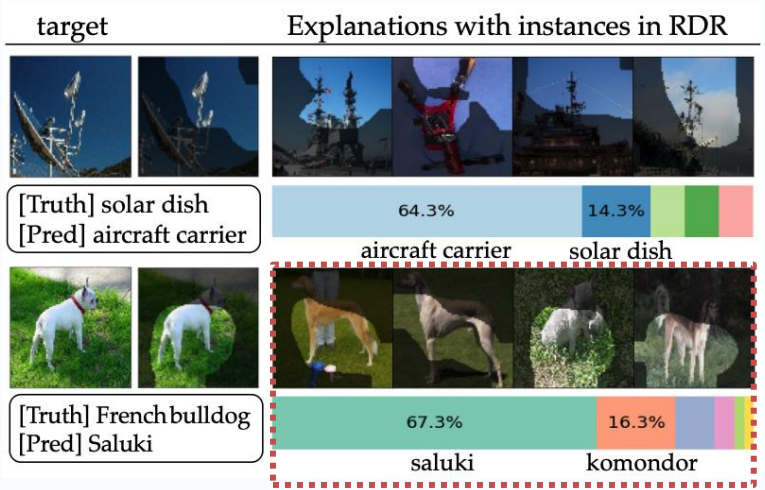
- Find unlabeled subclasses from data



Different RDRs capture different learned concepts without prior knowledge of sublabel information.

Misclassification Analysis

- Reasoning error-cases



The French Bulldog was misclassified since it has long and thin legs similar to Saluki.

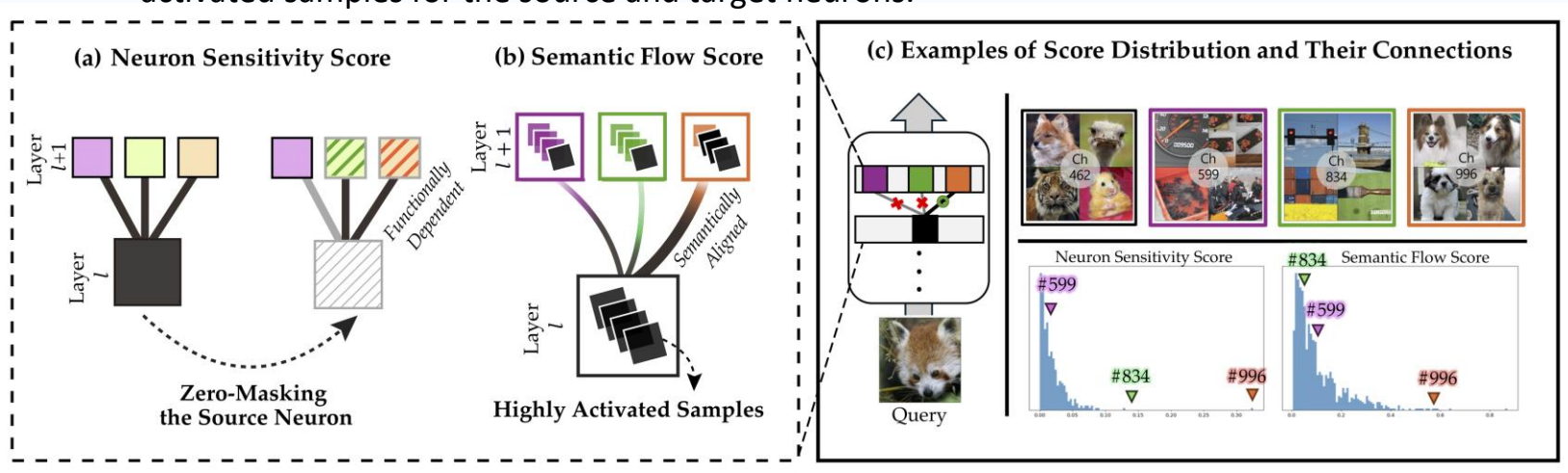
Wonjoon Chang, Dahee Kwon and Jaesik Choi, Understanding Distributed Representations of Concepts in Deep Neural Networks without Supervision, AAAI, 2024.

Question: How can we find concept-generating paths in DNNs?

- **From a black box to a decomposable blueprint**: fine-grained, query-specific concept circuit extraction
- Two score metrics:

  (1) **Neuron Sensitivity Score (S_{NS})** quantifies functional dependency by measuring how muting a source neuron impacts a target neuron's activation.

  (2) **Semantic Flow Score (S_{SF})** ensures semantic alignment by quantifying the overlap in highly-activated samples for the source and target neurons.



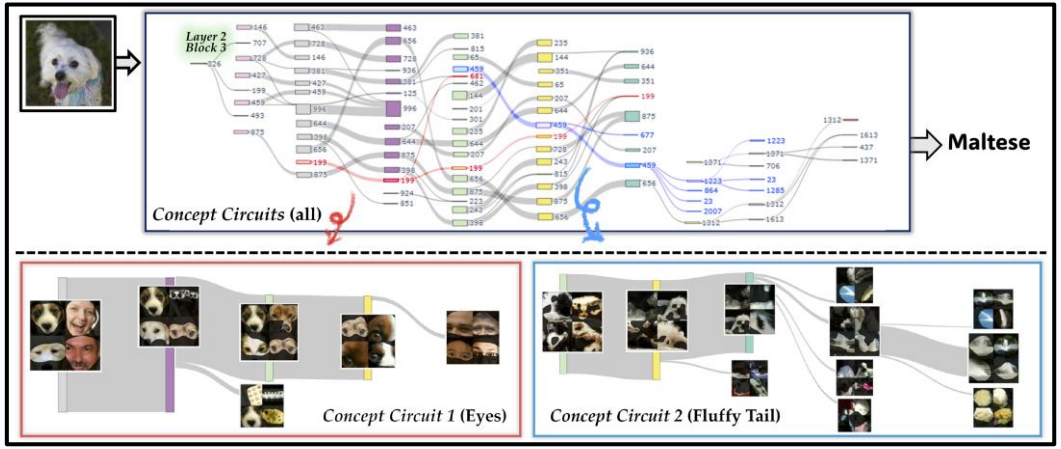
Dahee Kwon*, Sehyun Lee* and Jaesik Choi, **Granular Concept Circuits: Toward a Fine-Grained Circuit Discovery for Concept Representations**, **ICCV**, 2025.

Granular Concept Circuits: Toward a Fine-Grained Circuit Discovery [2025]

Question: How can we find concept-generating paths in DNNs?

From black box to a decomposable blueprint

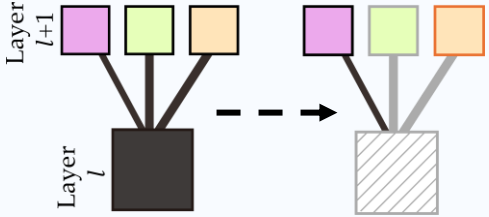
- Extract fine-grained, query-specific concept circuits
- Beyond circuit label



Two Score Metrics

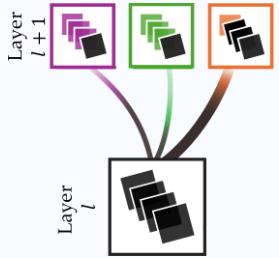
Neuron Sensitivity Score

Functional connectivity



Semantic Flow Score

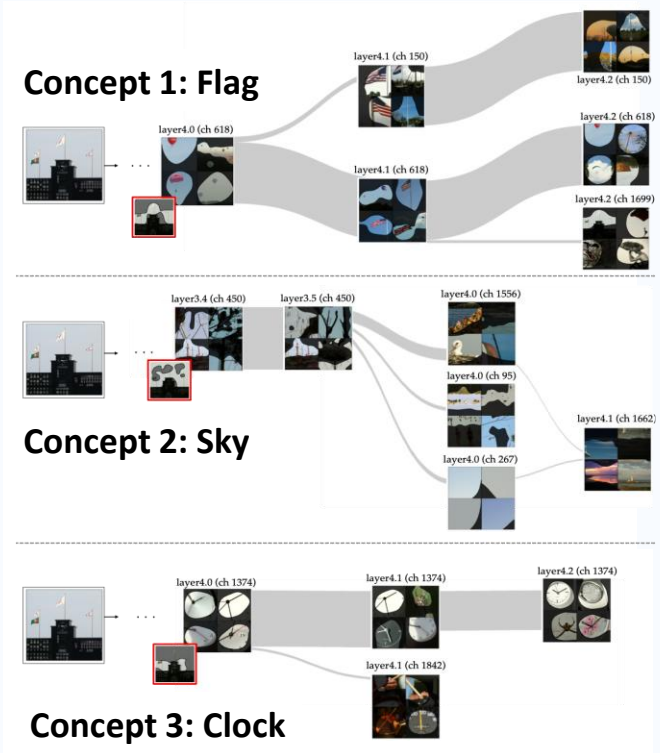
Interpretable connectivity



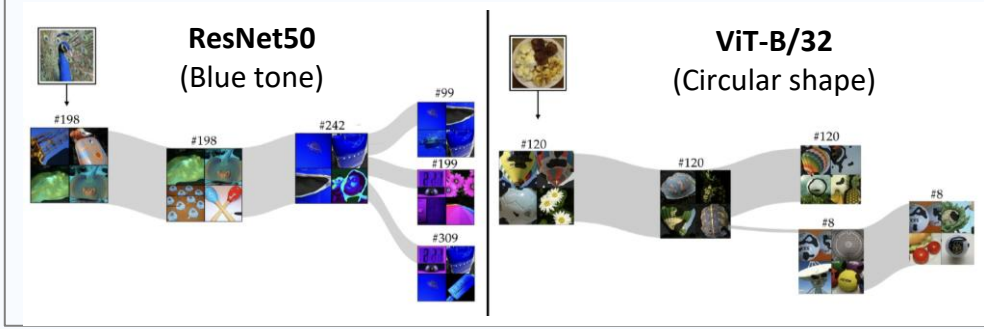
Dahee Kwon*, Sehyun Lee* and Jaesik Choi, **Granular Concept Circuits: Toward a Fine-Grained Circuit Discovery for Concept Representations**, **ICCV**, 2025.

Granular Concept Circuits: Toward a Fine-Grained Circuit Discovery [2025]

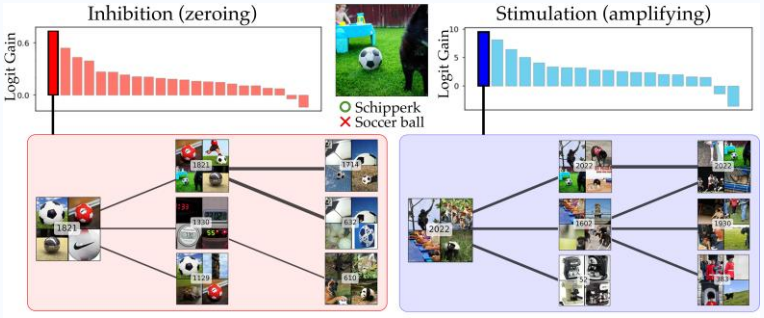
Multiple Concept Circuits from Single Query



Model-agnostic Method



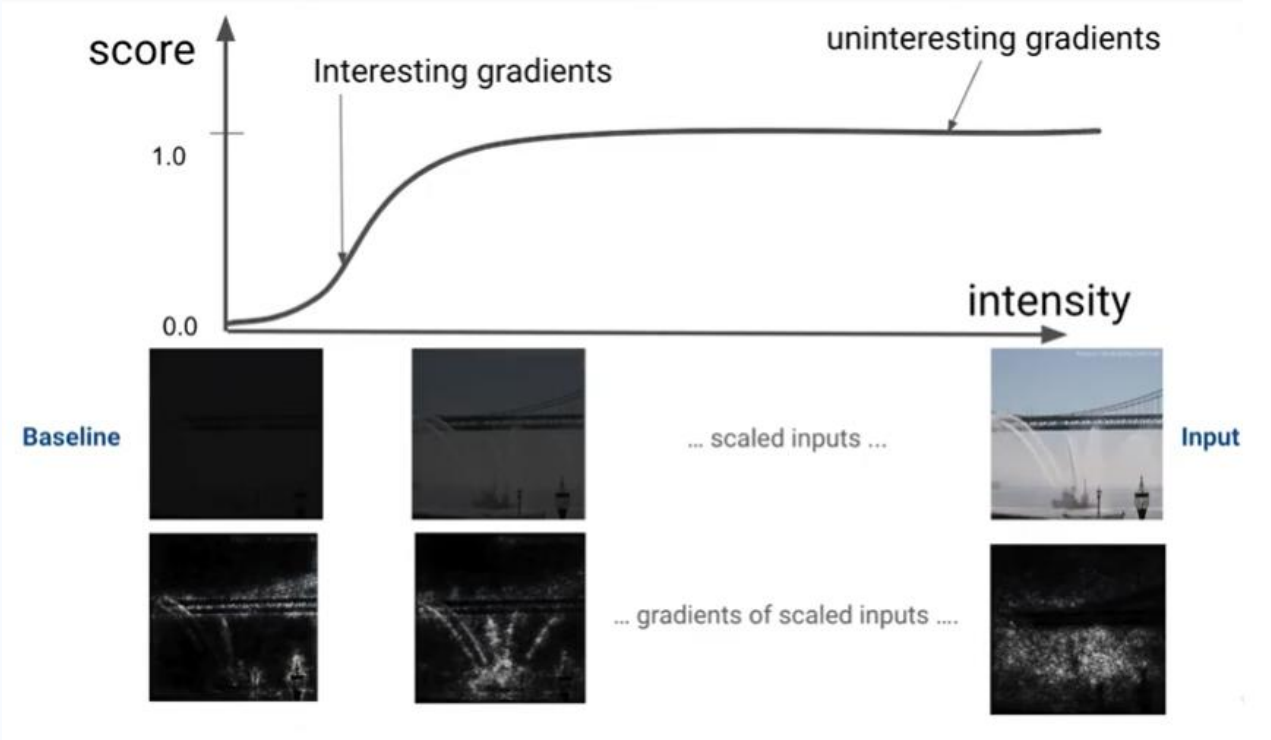
Auditing Misclassification



Dahee Kwon*, Sehyun Lee* and Jaesik Choi, **Granular Concept Circuits: Toward a Fine-Grained Circuit Discovery for Concept Representations**, **ICCV**, 2025.

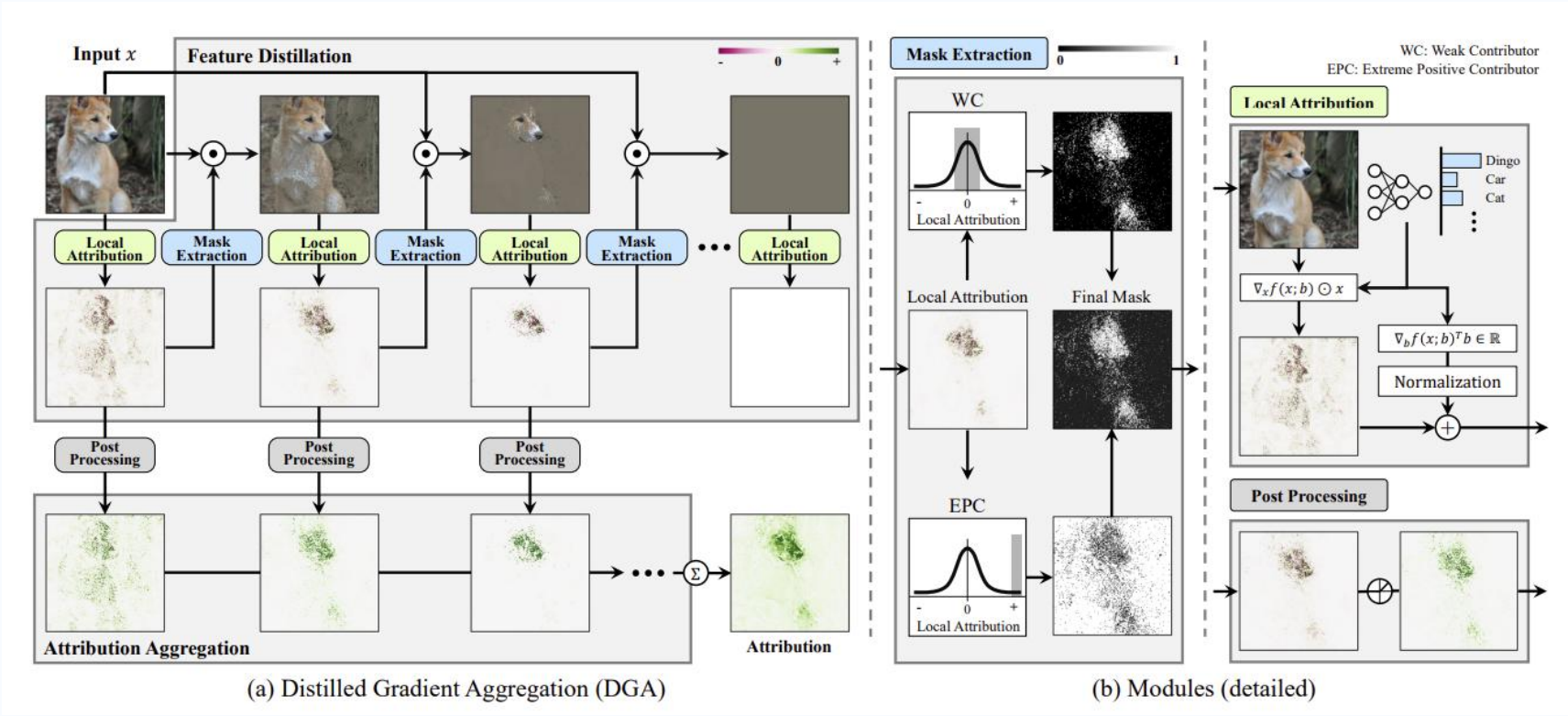
Question: How can we accurately compute input attributions of DNNs?

Input Attribution



Improved Input Attribution Method [2022]

Question: How can we accurately compute input attributions of DNNs?



Question: How can we accurately compute input attributions of DNNs?

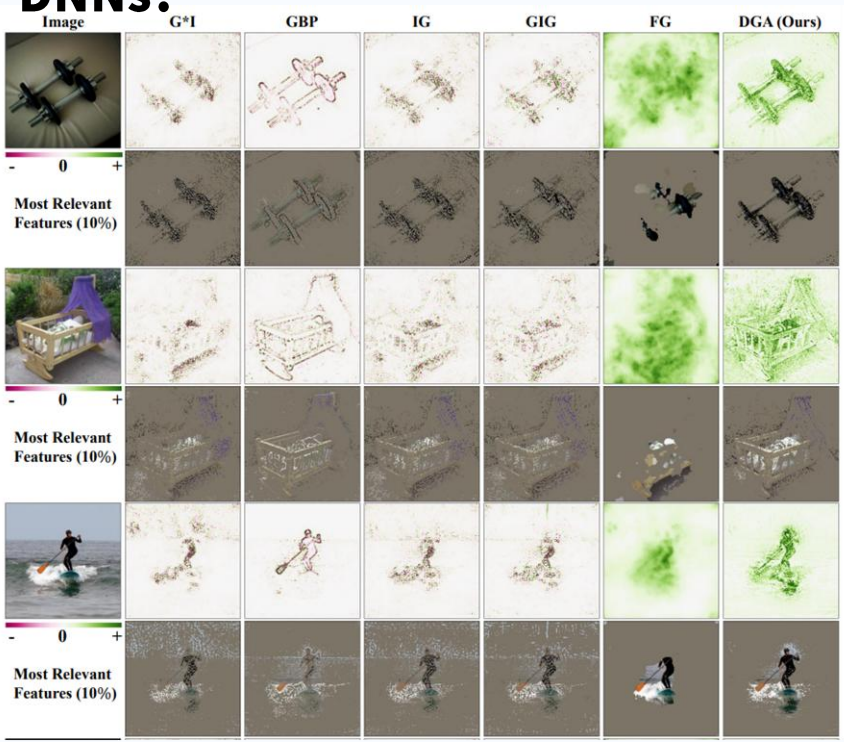


Table 1: Comparison of various attribution methods with LeRF and MoRF on three models.

		G*I	GBP	IG	FG	GIG	DGA
LeRF (↑ is better)	VGG-16	0.078	0.113	0.096	0.415	0.110	**0.434**
	ResNet-18	0.114	0.145	0.158	0.448	0.185	**0.533**
	Inception-V3	0.171	0.162	0.243	0.558	0.255	**0.691**
MoRF (↓ is better)	VGG-16	0.045	0.094	0.036	0.110	0.029	**0.023**
	ResNet-18	0.050	0.124	0.038	0.131	0.029	**0.019**
	Inception-V3	0.105	0.145	0.066	0.175	0.061	**0.041**

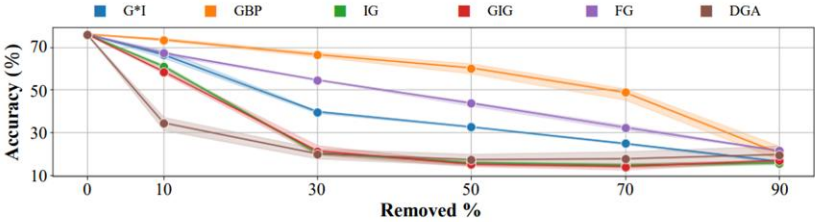


Figure 6: Comparison of ROAR experiment results on CIFAR-10 dataset among various attribution methods. The test accuracy for corresponding the percentage of removal.

Giyoung Jeon (INEEJI), Haedong Jeong, Jaesik Choi, Distilled Gradient Aggregation: Purify Features for Input Attribution in the Deep Neural Network, NeurIPS, 2022

Rethinking Shapley Value for Negative Interactions in Non-convex Games [2025]

Question: How should we compute input attributions when inputs are not independent?

Interactions in Shapley value

- The Shapley value is a main theoretic basis for a fair attribution rule, but its original formulation does not tell any interaction effects between features.

$$\phi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{1}{n} \binom{n-1}{s}^{-1} \Delta_i v(S) = \sum_{S \subseteq N \setminus \{i\}} \frac{1}{n} \binom{n-1}{s}^{-1} [v(S \cup \{i\}) - v(S)]$$

novel reformulation with **explicit interaction** terms

Theorem. *Shapley value is a weighted sum of interactions.*

$$\phi_i(v) = \Delta_i v(\emptyset) + \sum_{t=0}^{n-2} \frac{1}{n} \binom{n-1}{t}^{-1} \sum_{\substack{j \in N \\ j \neq i}} \sum_{\substack{T \subseteq N \setminus \{i,j\} \\ |T|=t}} I_{ij}(T)$$

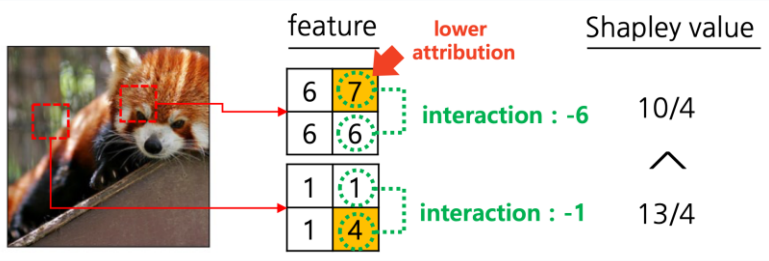
Interaction

$$I_{ij}(T) = \Delta_{ij} v(T) = \Delta_i v(T \cup \{j\}) - \Delta_i v(T)$$
$$= v(T \cup \{i,j\}) - v(T \cup \{i\}) - v(T \cup \{j\}) + v(T)$$

Undervaluation in Non-convex Games (ex. DNNs)

Cooperative behavior does not hold in non-convex games, where negative interactions arise: $I_{ij}(T) < 0$

→ conflict to Efficiency Axiom ($\sum_i \phi_i(v) = v(N)$)



Wonjoon Chang, Myeongjin Lee, Jaesik Choi, Rethinking Shapley Value for Negative Interactions in Non-convex Games, ICLR, 2025.

Rethinking Shapley Value for Negative Interactions in Non-convex Games [2025]

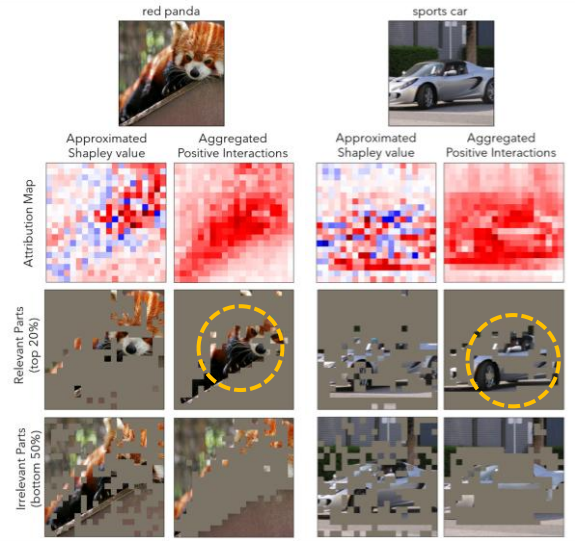
Question: How should we compute input attributions when inputs are not independent?

Aggregated Positive Interactions (API)

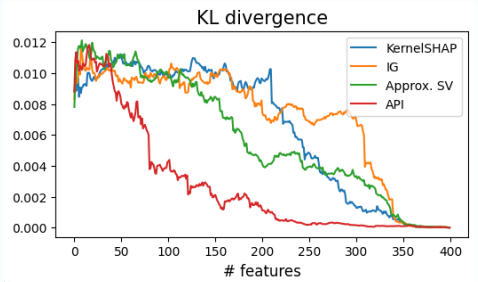
- Decompose each contribution into interactions and aggregates the positive parts, which represents the player's potential influence on improving the decision.

$$\phi_i(v) = \Delta_i v(\emptyset) + \sum_{t=0}^{n-2} \frac{1}{n} \binom{n-1}{t}^{-1} \sum_{\substack{j \in N \\ j \neq i}} \sum_{\substack{T \subseteq N \setminus \{i,j\} \\ |T|=t}} \max(I_{ij}(T), 0)$$

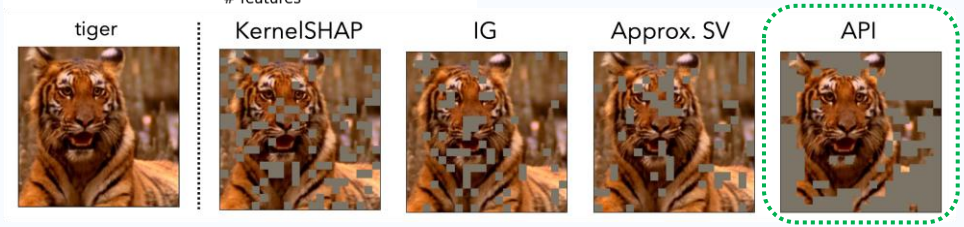
Attribution Result



Quantitative Result



Evaluate the change in output as highly attributed features are added first.
→ With API attribution, we need much smaller number of features to recover original decision.
→ more compact explanation!

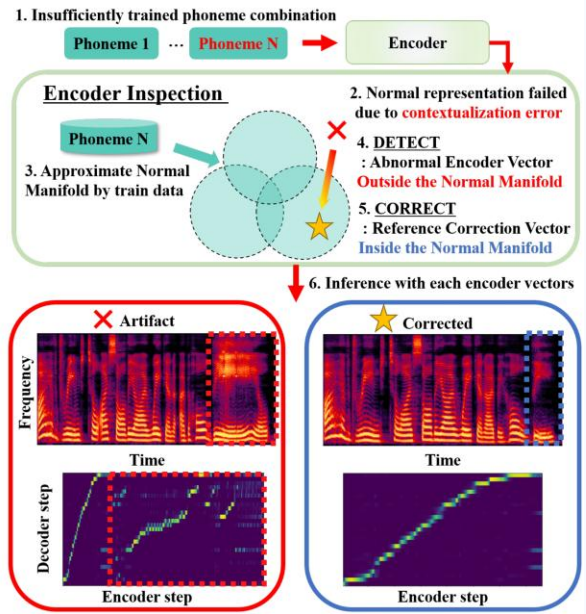


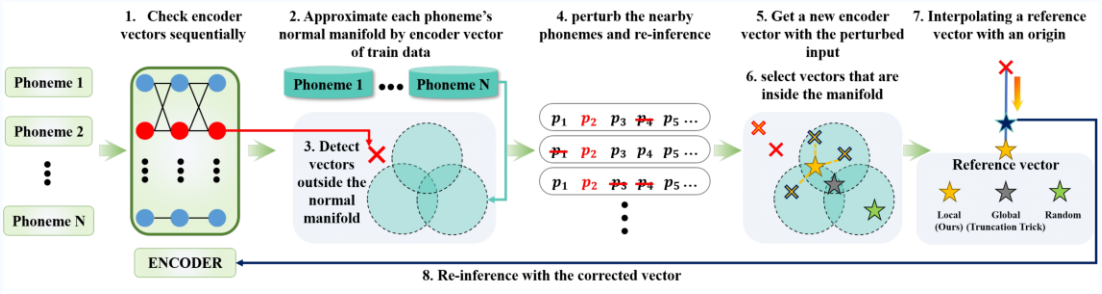
Wonjoon Chang, Myeongjin Lee, Jaesik Choi, Rethinking Shapley Value for Negative Interactions in Non-convex Games, ICLR, 2025.

Automatic Corrections of Artifacts in Neural Text-to-Speech Models [2025]

Question: Can we fix errors in speech generation models?

- **Contextualisation errors** within a text-to-speech (TTS) model due to insufficient text context learning
- Understand how model generate contextualisation errors and proposed methods to **automatically detect and correct errors**





Algorithm 1 Detection of abnormal context vectors

Parameters l: encoder layer, n: input index, k: K-nearest neighbors algorithm parameter, p: phoneme type, ϕ_r: encoder vector of train sample, ϕ_g: encoder vector of test sample, $\mathbf{\Phi_r}$: encoder vectors of overall train data, $e()$: encoder forward function, $\mathrm{NN}_k(\phi_r, \mathbf{\Phi_r})$: kth nearest neighbor

$f(\phi_g, \mathbf{\Phi_r}) : \|\phi_g - \phi_r\|_2 \leq \|\phi_r - \mathrm{NN}_k(\phi_r, \mathbf{\Phi_r})\|_2$ (for at least one $\phi_r \in \mathbf{\Phi_r}$)
for $l = 1$ to L **do**
  $\phi^l_{g,1:N} = e^{l-1:l}(\phi^{l-1}_{g,1:N})$
  for $n = 1$ to N **do**
    if $f(\phi^{l,p}_{g,n}, \mathbf{\Phi^p_r})$ **then**
      $\phi^{l,p}_{g,n}$ is inside of manifold
    else
      $\phi^{l,p}_{g,n}$ is outside of manifold
    end if
  end for
end for

Algorithm 2 Correction of abnormal context vectors

Parameters l: encoder layer, l_{target}: encoder layer to be corrected, n: input index,$\mathbf{\Phi_r}$: encoder vectors of the overall training data, ϕ_g: encoder vector of the test sample, $e()$: encoder forward function, p_n:nth input phoneme, and ψ:truncation trick parameter

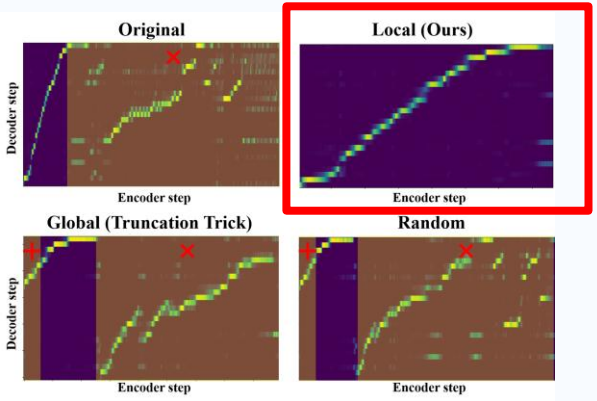
for $l = 1$ to L **do**
  $\phi^l_{g,1:N} = f^{l-1:l}(\phi^{l-1}_{g,1:N})$
  if $l = l_{target}$ **then**
    for $n = 1$ to N **do**
      if $\phi^l_{g,n}$ is outside of manifold **then**
        $\boldsymbol{P} \leftarrow$ Perturb the inputs around p_n
        $\boldsymbol{\phi}^l_{g,n} \leftarrow e^{1:l}(\boldsymbol{P})$
        $\boldsymbol{\phi}^{in,l}_{g,n} \leftarrow$ Detect normal of $\boldsymbol{\phi}^l_{g,n}$
        $\overline{\phi}^l_{g,n} \leftarrow Mean(\boldsymbol{\phi}^{in,l}_{g,n})$
        $\phi^l_{g,n} \leftarrow \overline{\phi}^l_{g,n} + \psi(\phi^l_{g,n} - \overline{\phi}^l_{g,n})$
      end if
    end for
  end if
end for

Seongyeop Jeong, June Sig Sung, Inchul Hwang and Jaesik Choi, An Automated Method to Correct Artifacts in Neural Text-to-Speech Models, T-ASLP, 2025

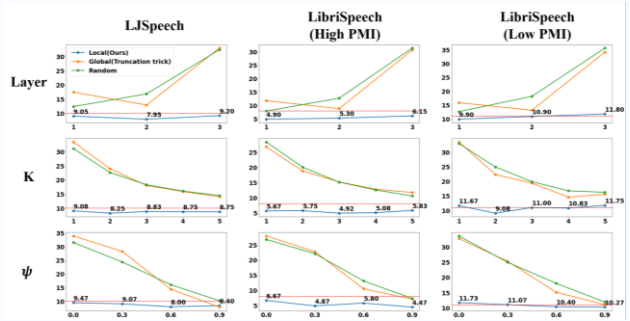
Automatic Corrections of Artifacts in Neural Text-to-Speech Models [2025]

Question: Can we fix errors in speech generation models?

Alignment error between text and speech

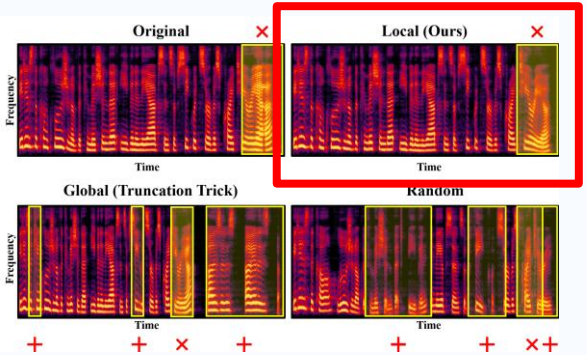


Alignment error between text and speech (↓)

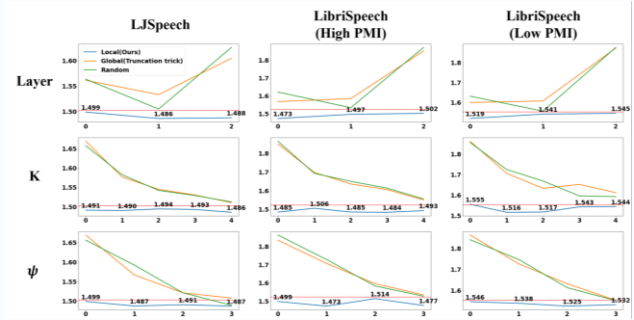


Local(Ours)
Global(Truncation trick)
Random

Mel-spectrogram Quality



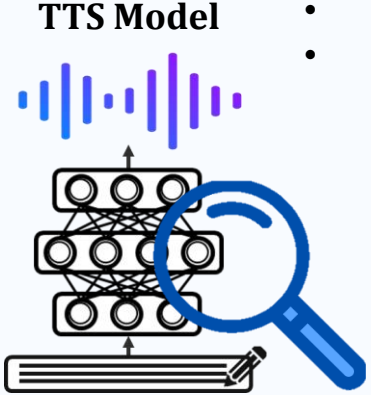
Frechet Wav2Vec distance (↓)



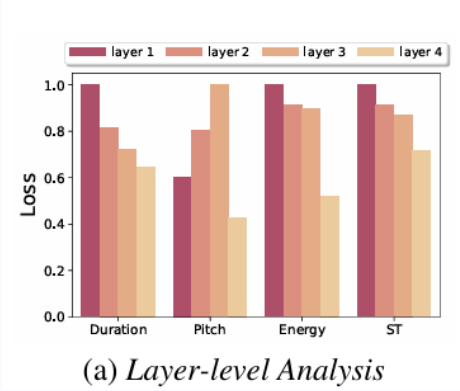
Seongyeop Jeong, June Sig Sung, Inchul Hwang and Jaesik Choi, An Automated Method to Correct Artifacts in Neural Text-to-Speech Models, T-ASLP, 2025

Post-hoc Prosody and Mispronunciation Corrections in TTS Models [2025]

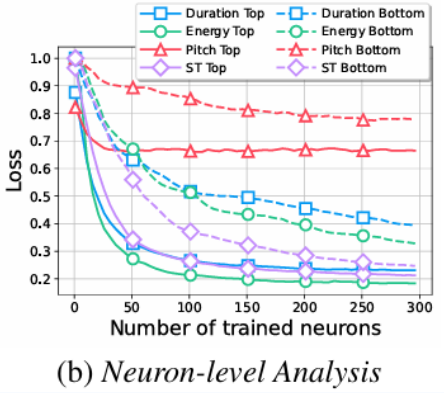
Question: Can we fix errors in speech generation models?



- Q1: Do the internal activations of a Tacotron2 encoder contain acoustic information?
- A1: Yes

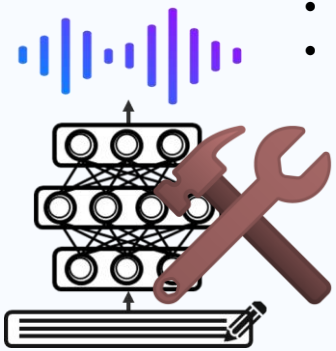


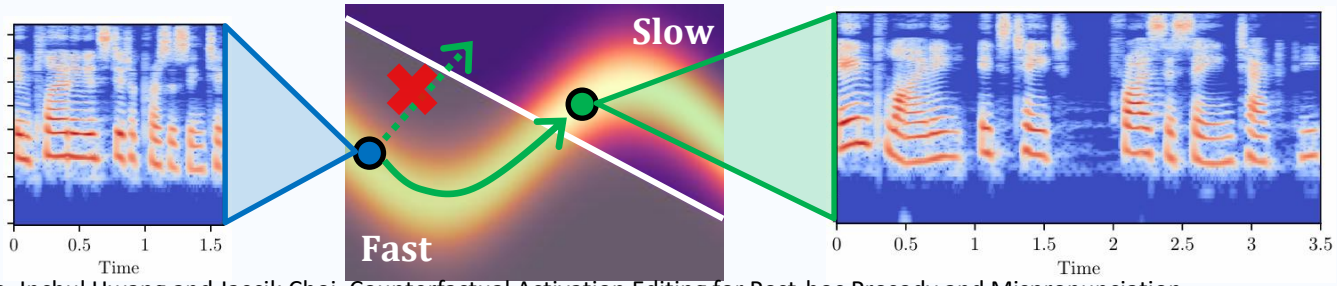
(a) *Layer-level Analysis*



(b) *Neuron-level Analysis*

- Q2: How can we edit encoder activations to manipulate prosody?
- A2: Edit activations along the manifold



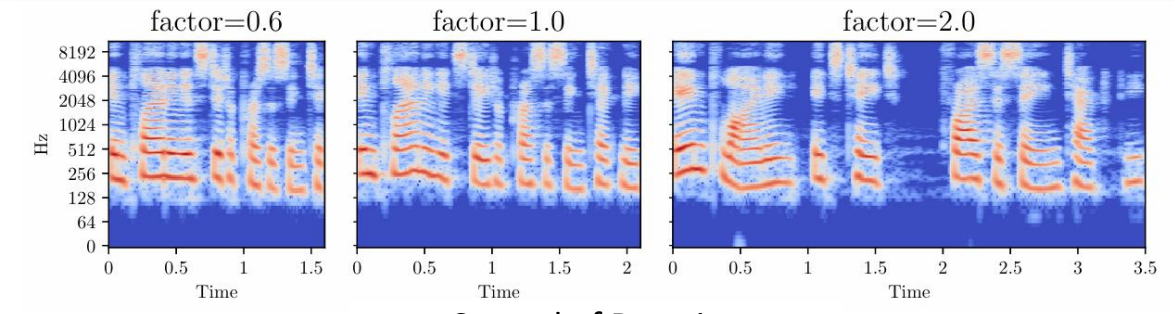


Kyowoon Lee, Artyom Stitsyuk, Gunu Jho, Inchul Hwang and Jaesik Choi, Counterfactual Activation Editing for Post-hoc Prosody and Mispronunciation Correction in TTS Models, Interspeech, 2025.

Post-hoc Prosody and Mispronunciation Corrections in TTS Models [2025]

Question: Can we correct the errors of speech generation models?

- Post-hoc control of prosody without retraining model.



Kyowoon Lee, Artyom Stitsyuk, Gunu Jho, Inchul Hwang and Jaesik Choi, Counterfactual Activation Editing for Post-hoc Prosody and Mispronunciation Correction in TTS Models, Interspeech, 2025.

Real-Time Explainable AI for Acute Kidney Injury Prediction

1. EMR
2. Preprocessing
3. AI-training logic
4. Clinical Insights
5. Department-specific preprocessing
6. Performance

KAIST – SNUBH AKI Prediction Project

- **Developing and clinically applying an AI system that (1) predicts AKI within 48 hours and (2) explains the reasons.**

Definition of EMR data for AKI prediction

SNUH 분당서울대학교병원

Patient Information

Basic information						Underlying disease	Prescribing medication before admission
Age	Sex	BMI	ICU	Base Cr	Base eGFR	19 underlying disease + Charlson Comorbidity Index	16 prescription information

Others: smoking, drinking, surgery

Time Series Data after admission

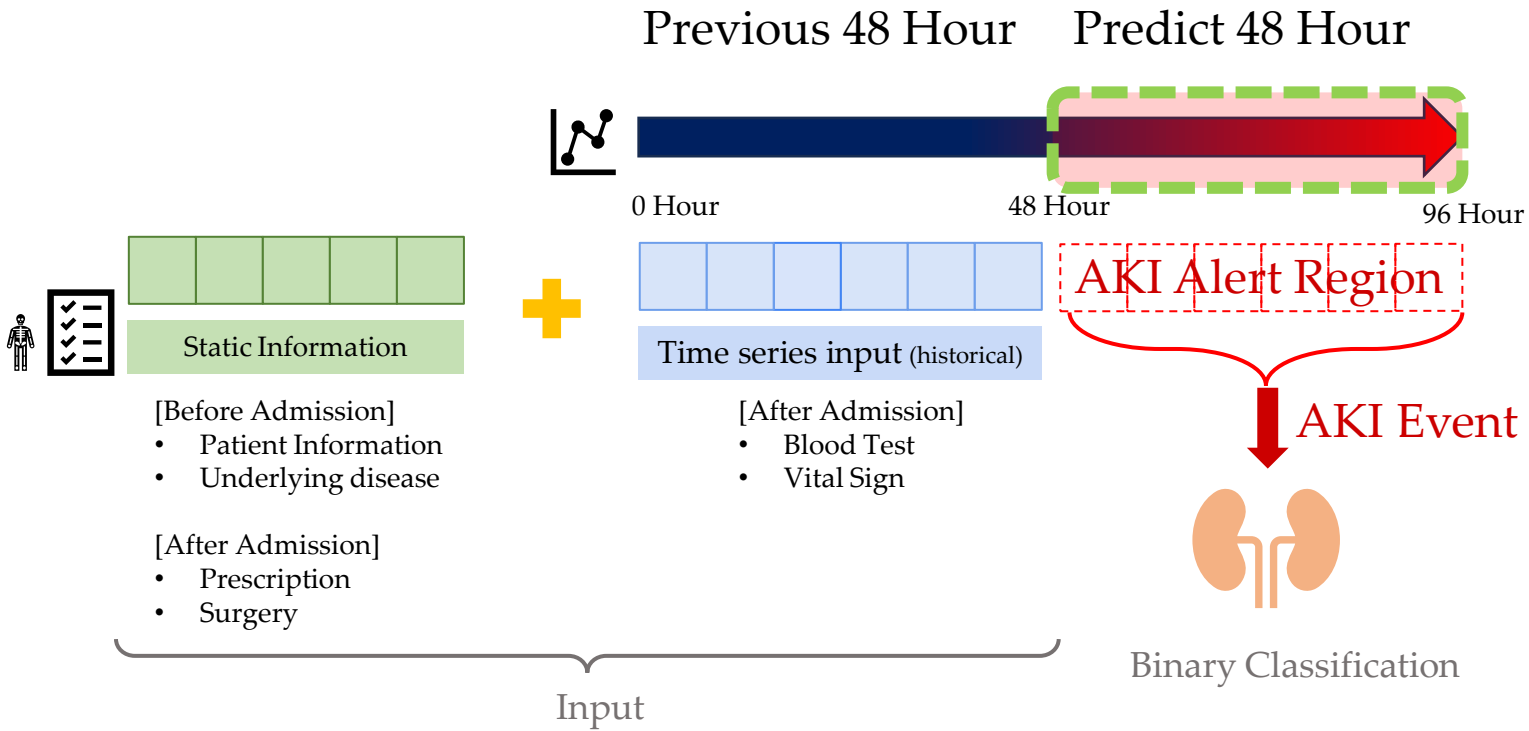
Prescribing medication (Per Day)	Surgery & anesthesia & ICU (Per Day)			Blood/Urine Test (Per Day & Time)	Vital signs (Per Day & Time)			
Prescription information	Major/minor surgery	General anesthesia	Surgical times	Values for each test	SBP	DBP	BPM	Body Temperature

avg / low / high / max / min

AKI occurrence within next 48h

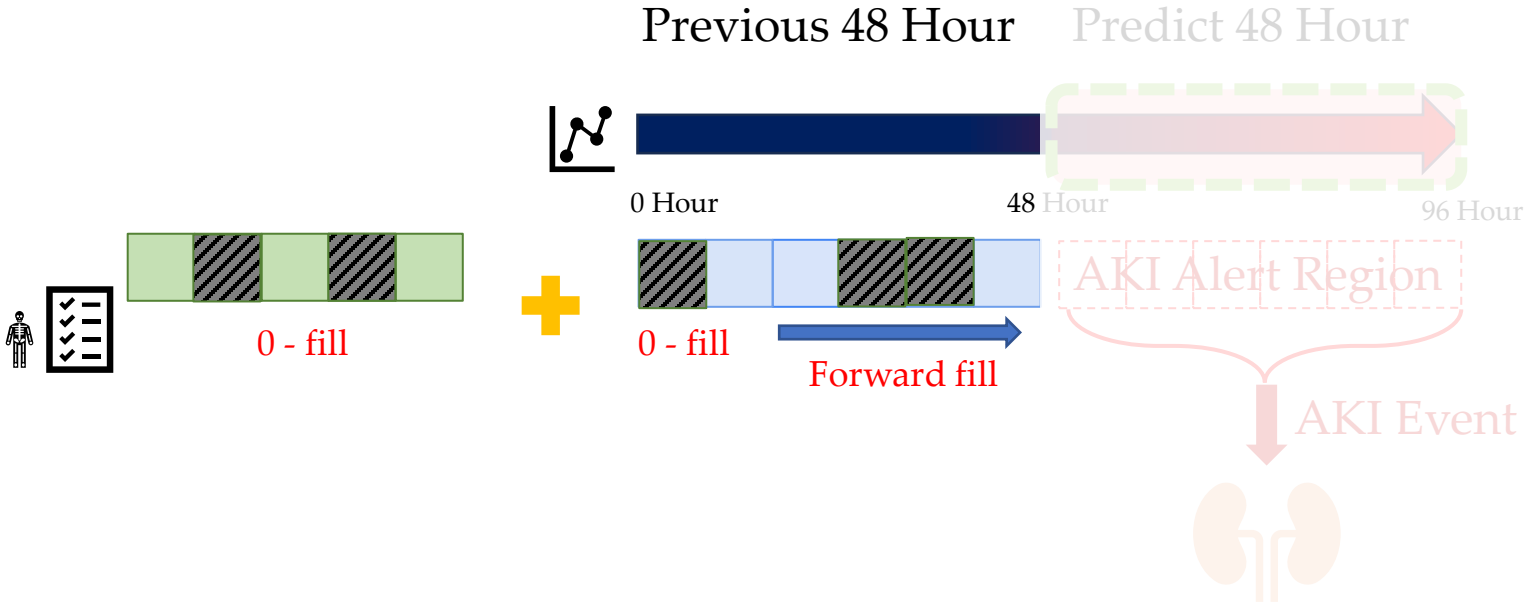
AKI Occurence (Per Day & Time)
Any AKI Occurence

EMR Data Preprocessing for AKI Prediction



EMR Data Preprocessing for AKI Prediction

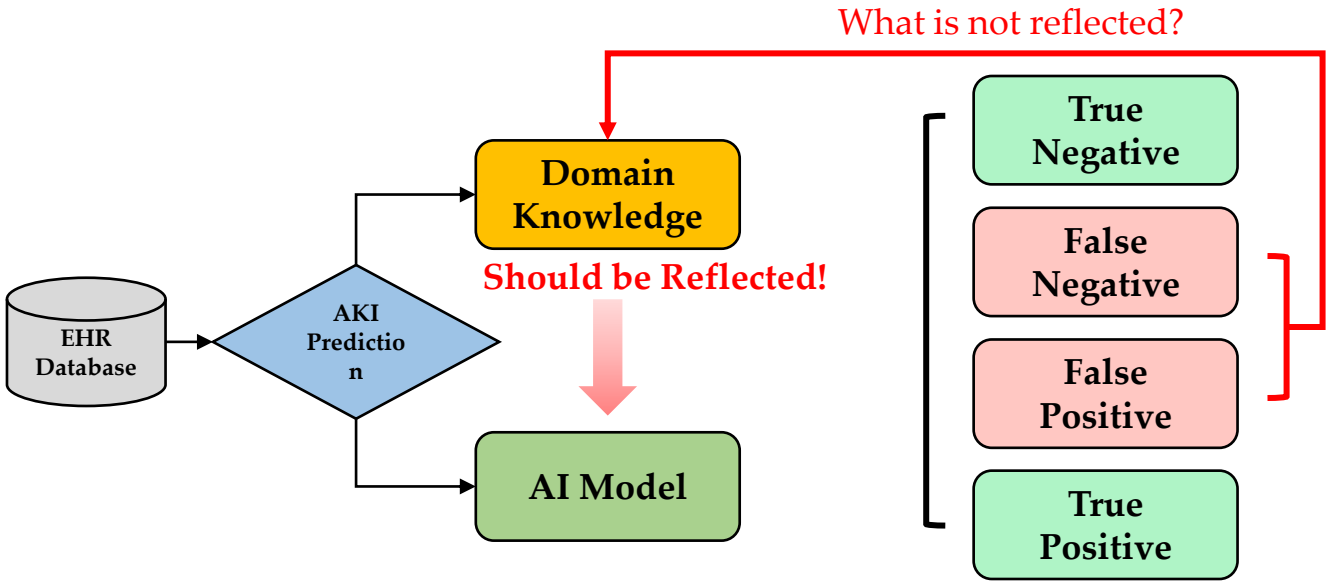
- **Depending on data-type:**
  - **Forward-fill for time-series trends; Zero-fill for missing tests/meds.**
  - **For missing test values or absent medication use, apply zero-fill to preserve information**



Null Value (결측값)

AI-training Logic

- **Analyzing learned vs. missing AKI knowledge → Infusing clinical insights for model improvement.**



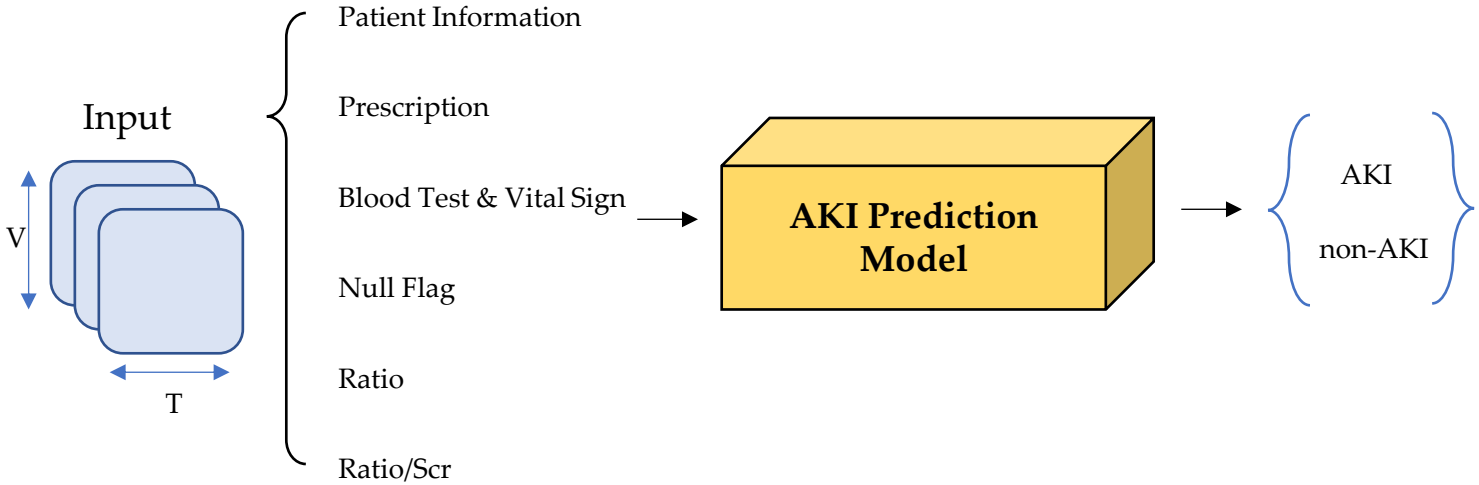
Clinical Insights to Data Preprocessing

- **Added a binary variable indicating the presence or absence of side effects from AKI-related medications.**
- **Example: Vasopressors may cause side effects raising blood pressure.**

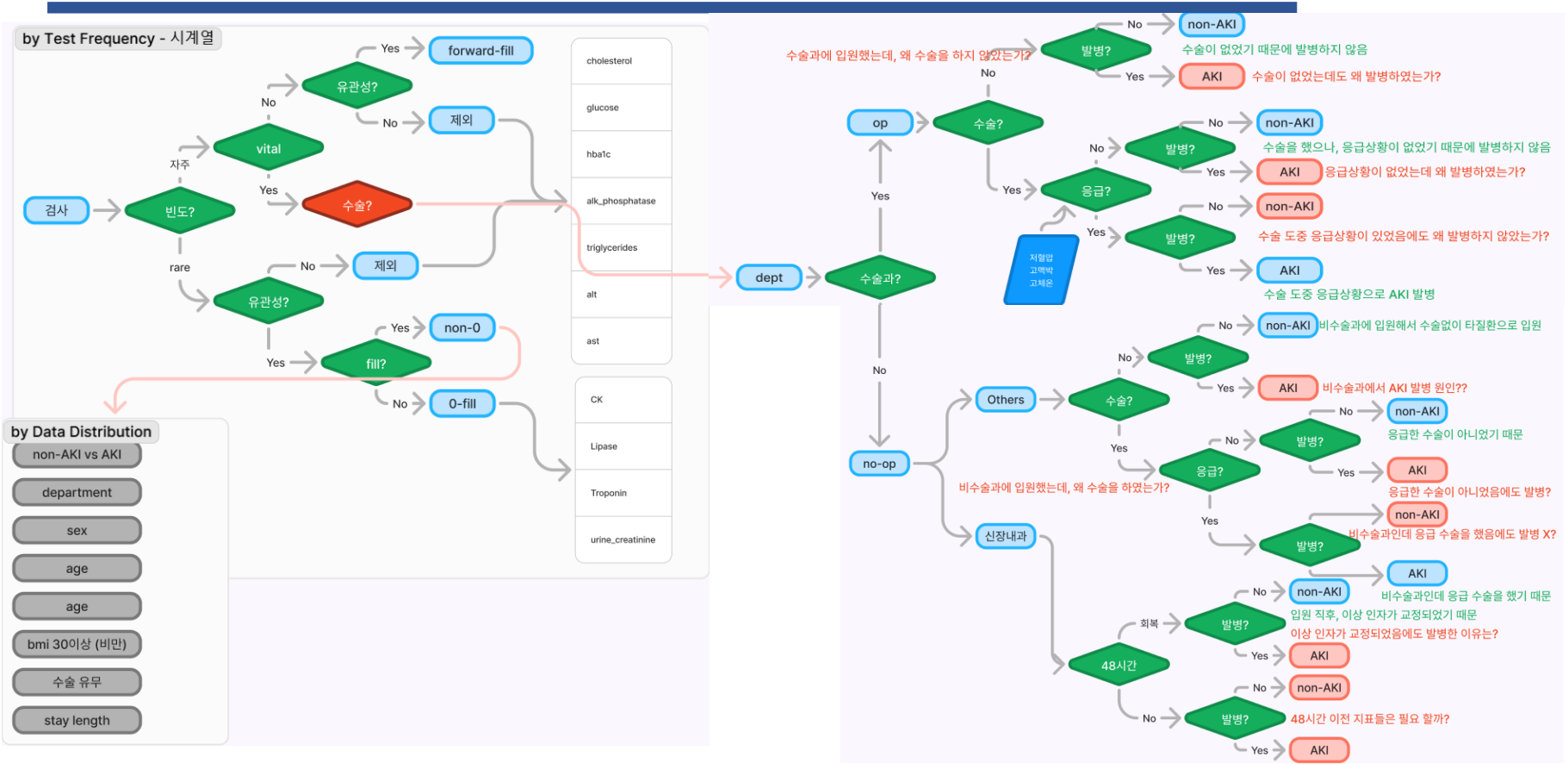
	Renal Function		Vital Signs			Drug-induced AKI							
			Blood Pressure		Pulse Rate	Hemodynamically-mediated injury	Inflammation	Renal Hypoperfusion	Rhabdomyolysis	Thrombotic microangiopathy	Crystal nephropathy	Acute Tubular Necrosis	Tubular Injury
Medicines	Nephrotoxicity	Relevant	Up	Down	Relevant								
arb	☐	☑	☐	☑	☑	☑	☐	☐	☐	☐	☐	☐	☐
betablocker	☐	☑	☐	☑	☑	☐	☐	☐	☐	☐	☐	☐	☐
acei	☐	☑	☐	☑	☑	☑	☐	☑	☐	☐	☐	☐	☐
acyclovir	☐	☑	☐	☐	☐	☐	☐	☐	☐	☐	☑	☐	☐
aminoglycoside	☑	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☑	☑
amphotericin	☑	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☑	☑
ccb	☐	☑	☐	☑	☑	☐	☐	☐	☐	☐	☐	☐	☐
cisplatin	☑	☐	☐	☐	☑	☐	☐	☐	☐	☐	☐	☑	☑
cyclosporine	☑	☐	☐	☐	☑	☑	☐	☑	☐	☑	☐	☐	☑
diuretics	☐	☑	☐	☐	☐	☑	☐	☐	☐	☐	☐	☐	☐
nsaid	☑	☐	☐	☐	☑	☑	☑	☑	☐	☐	☐	☐	☑
tacrolimus	☑	☐	☐	☐	☑	☑	☐	☑	☐	☑	☐	☐	☐
vancomycin	☐	☑	☐	☐	☐	☐	☑	☐	☐	☐	☐	☐	☑
vasopressor	☐	☑	☑	☐	☑	☐	☐	☐	☐	☐	☐	☐	☐
colistin	☑	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☑
statin	☐	☑	☐	☐	☑	☑	☐	☐	☑	☐	☐	☐	☑

Final Data-Frame for AKI Prediction

- **To learn temporal patterns, the dataset was reformatted into (Variable, Time) frames.**
- **Model design with separate embeddings for different feature types**

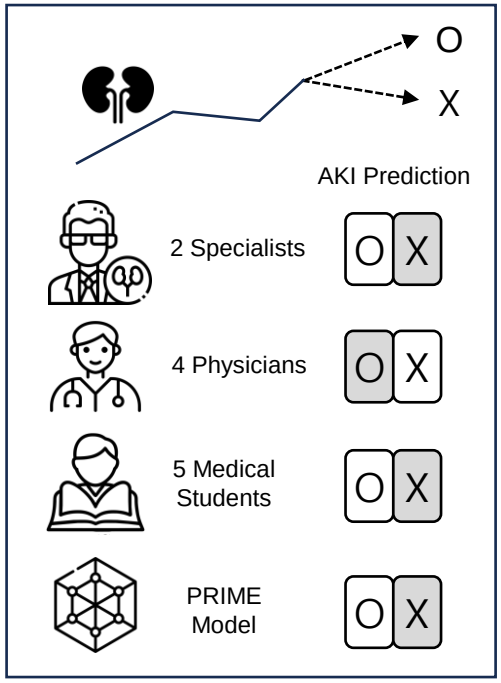


Department-specific preprocessing

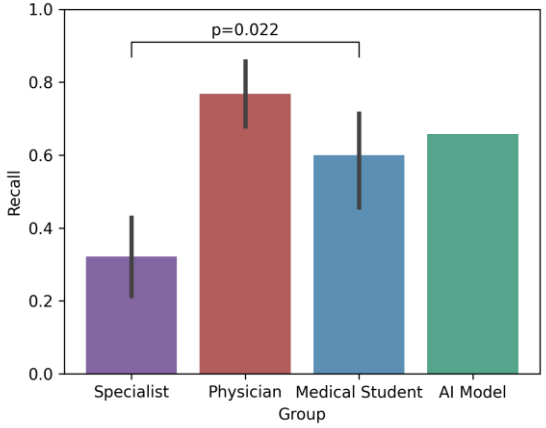


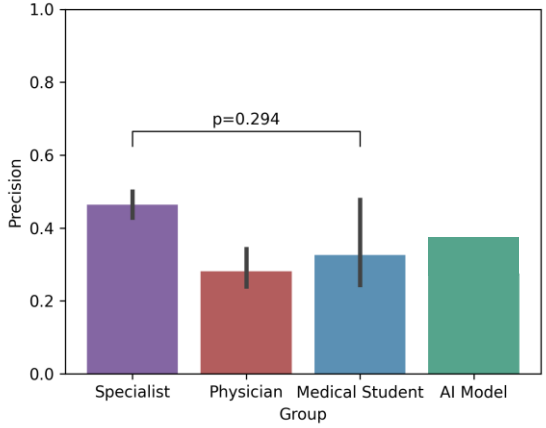
AI모델과 임상의

- 임상의와 AI모델의 성능 비교



	Recall	Precision
Specialist	32% (4)	47% (1)
Physician	77% (1)	28% (4)
Medical Student	60% (3)	33% (3)
AI(ours)	71% (2)	38% (2)

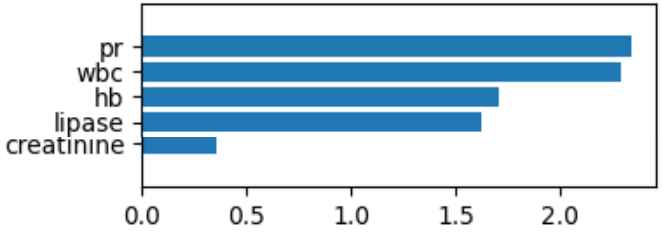




Meta-Explainability: A New Paradigm of Explainability in Medical AI

- Conventional explainability reduces interpretability by only ranking input importance.
- Adding 'explainability' to explainability: user-friendly and domain-specific medical explanations.

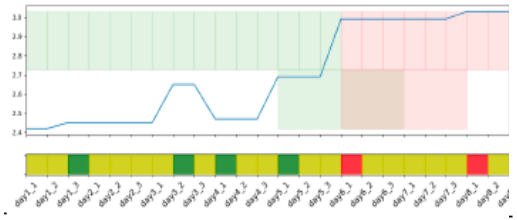
[Traditional Approach]
Feature Ranking (Top 5)



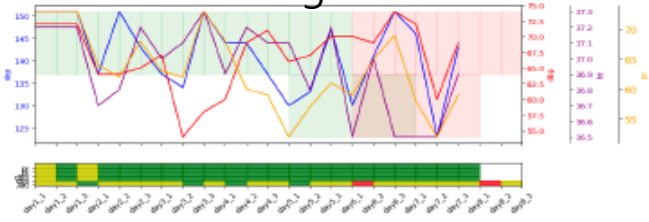
[New Paradigm – Meta Explanation]

- Creatinine is important because it shows abnormal values.
- Vital signs matter due to sharp changes from previous readings.

Creatinine



Vital Sign



PR (pulse rate), WBC (white blood cell), HB (hemoglobin), Lipase

** This data is synthesized and anonymized.*

Research Questions in XAI

Q1: Can we find the role of internal units in Large Language Models (LLMs)?

Q2: Can we find internal units which behaves badly (e.g., making artifacts) in LLMs?

Q3: Can we control internal units which behaves badly in LLMs?

Q4: Can we correct internal units which behaves badly in LLMs in an unsupervised way?

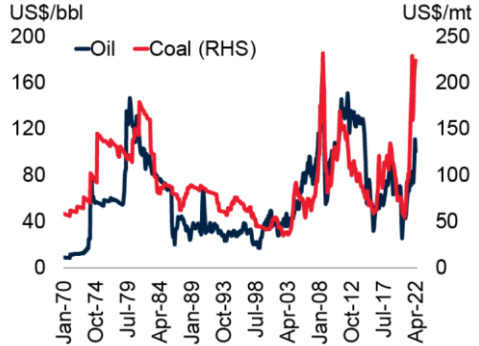
Q6: Can we improve the input attribution method by finding the better path in LLMs?

Thank you

jaesik.choi@kaist.ac.kr
jaesik@ineeji.com

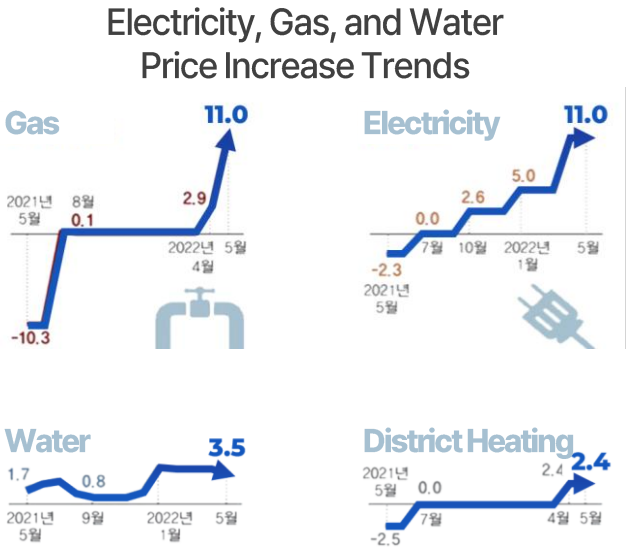
INEEJI Inc. – Domestic Application Case

Current Situation: Rising Energy Costs



Source: World Bank.

Note: Monthly data from 1970 to April 2022. Prices deflated by the U.S. Consumer Price Index.



Raw Material & Energy Costs

(출처: JUSTIN-DAMIEN GUÉNETTEJEETENDRA KHADAN, World Bank Blogs, "The energy shock could sap global growth for years", 2022.06.22 | 통계청)

INEEJI INFINITE OPTIMAL SERIES™

AI Solutions for Process Efficiency and Energy Cost Reduction

Core engine

AI-Based Prediction	AI-Driven Automatic Control	Raw Material Price, Sales Price, and Demand Consideration
PREDICT	**EXPLAIN**	**COSTSAVER**
		
AI for Production Process Prediction and Optimization	AI Explaining the Reasons for Process Optimization	AI for Improving Product Spread
AI Time-Series Forecasting for Process Optimization **State-of-the-Art**	**Explainable AI Process Explanation Solution** First Visualization of Time-Series Deep Learning Decisions	AI Solution for Short- and Long-Term Raw Material Price Forecasting **Maximize Production Profit**

Core technology

Automated Data Preprocessing Technology	Causal Relationship Extraction in Key Manufacturing Predictors	Multivariate Time-Series Forecasting for Processes	Time-Series Forecasting Model Optimization Technology	XAI for Equipment Prediction and Process Optimization

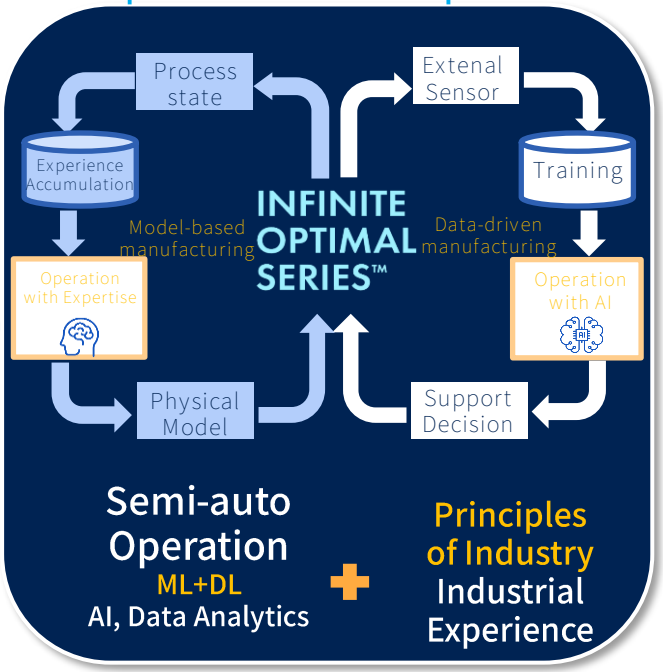
INEEJI INFINITE OPTIMAL SERIES™

Manufacturing Site

Sensor Data
(current, vibration, distortion surface temperature, etc.)

데이터 수집
(Analog Digital, LAN)

Operator expertise → INEEJI → AI automatic operation



Semi-auto Operation
ML+DL
AI, Data Analytics

+

Principles of Industry
Industrial Experience

- Improved productivity & reduced lead time
- Reduction of raw material & energy costs
- AI-based flexible process operation

INEEJI INFINITE OPTIMAL SERIES™

Key Sectors in Manufacturing and Industry

Steel: AI Heat Control for Blast\Electric\Reheating Furnaces

Cement: AI Heat Control for Preheater and Kil

Petrochemicals: Reaction Prediction & Optimization AI

Power Plant: Boiler Optimization AI

지속 가능 및 사업 다각화 사이트

- Blast Furnace
- Glass Melting Furnace
- Cathode Calcination Process
- Semiconductor Assembly Process
- Bio/Pharmaceuticals
- Wind/Hydropower Plant

INEEJI Inc. – Domestic Application Case

Energy Efficiency

posco

조선일보 / 2021년
인이지 창업 전
UNIST/KAIST 협업

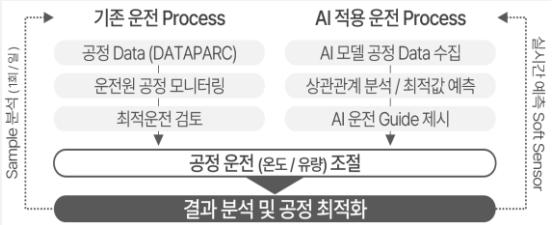
- Smart Blast Furnace with AI for real-time monitoring & control
- +240 tons/day molten iron (+5%), 1% fuel cost reduction
- WEF Lighthouse Factory, Pohang (Jul 2022)

Top 5 use cases	Impact
Machine vision and deep learning	↑ 4% Production output
Visualization and digitalization	↑ NA Production output
AI-based BOF temperature control	↓ NA Cost
Machine learning for rolling force	↑ 5% Productivity
AI-based automatic control	↓ 60% Quality deviations

SK picglobal

한경 AI 미래포럼 / 2022년

- AI-driven process optimization cuts energy costs without extra
- +2 tons/day PO production, 400M KRW annual savings
- AI·DT seen as survival, not choice



KG Steel

철강금속신문 / 2023년

- Industry-first AI predictive heat-treatment control in CGL (Continuous Galvanizing Line)
- Real-time process data collection & 2-hour-ahead quality prediction/optimization
- ~98% prediction accuracy, improved quality, and ₩450M annual energy savings

(출처: World Economic Forum in collaboration with McKinsey&Company, "Global Lighthouse Network:Insights from the Forefront of the Fourth Industrial Revolution", 2019.12.)

INEEJI Inc. – Domestic Application Case

Contents	Energy-saving Effect		Energy Source
	Improv.(%)	TOE/1year	
Steel heat-treatment property prediction & optimal temperature control	3	149.3(1)	LNG
Glass furnace temperature prediction with optimized control of fuel input and electric booster heating.	3	218.6(2)	LNG
Preheater temperature prediction and optimized fuel input control (pulverized coal, recycled fuel) in a rotary kiln	5	1,540(3)	Pulverized coal
Smart intersection traffic signal optimization with congestion prediction for about 200 vehicles	7	356(4)	Gasoline
RHDS diesel quality prediction (within 0.8% error) & fuel consumption optimization	1	261.3(5)	by-product fuel
Electric arc furnace operation optimization through molten time prediction using scrap composition and weight	7.1	954.9(6)	Electricity
Gas/steam turbine maximum output prediction in combined cycle power plants based on equipment condition	-	-	-
AI-based additive dosage prediction and control technology development	-	-	-

1. LNG savings: 2.79 nm^3/min → 1,466,424 nm^3/year → 149.3 TOE/year
2. B-C oil savings: 3% per ton → 600 L/day at 250 t/day → 218.6 TOE/year
3. Coal savings: 0.3 t/h → 1,758,000 kcal/h → 1,540 TOE/year
4. CO_2 reduction: 1,000 t/year, equivalent to 460,000 L gasoline → 356.5 TOE/year
5. By-product fuel oil savings: 33.5 L/h → 293,626 L/year → 261.3 TOE/year
6. Power savings: 476 kWh/heat → 11,424 kWh/day → 4,169,760 kWh/year → 954.9 TOE/year

INEEJI INFINITE OPTIMAL SERIES™ Use Case

Carried out by INEEJI

Best Practice Case

KG Steel

철강금속신문 / 2023

KG Steel has launched the industry's first AI-driven CGL furnace heat treatment prediction and control system.

The AI updates every 15 seconds, **predicts product quality up to 2 hours ahead**, and maintains optimal operations.

It **achieves 98% prediction accuracy**, ensuring quality competitiveness, while **reducing LNG** consumption for significant cost savings.

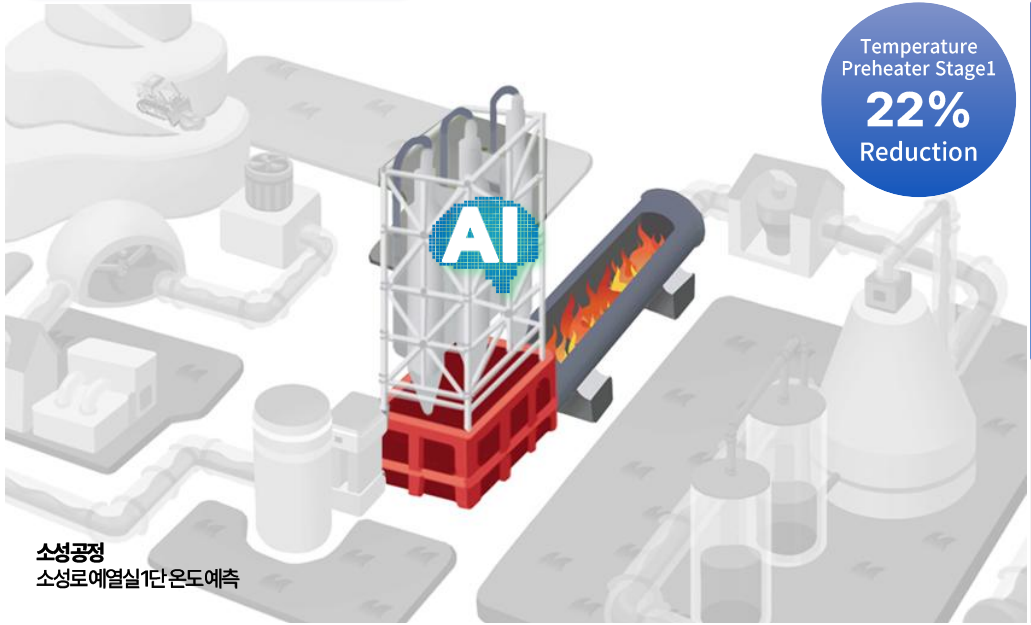
KG스틸 당진공장 연속용융아연도금(CGL) 운전실에서 작업자가 'AI 열처리 예측제어 시스템'을 모니터링하는 모습./ KG스틸 제공

INEEJI INFINITE OPTIMAL SERIES™ Use Case

Optimization of Alternative Fuel Use in Cement Kilns

INFINITE OPTIMAL SERIES™ THINK SMART

Carried out by INEEJI



소성공정
소성로예열실1단온도예측

As-is	Excess CO_2 from coal combustion Quality issues from fuel fluctuations Challenge in optimizing alternative fuel use
To-be	3% ▼ in coal use 22% ▼ in Preheater Stage 1 temperature Increased alternative fuel ratio Reduced carbon emissions

- To reduce carbon emissions from coal combustion, alternative fuels were introduced but caused unstable calorific fluctuations, limiting substitution.
- By predicting unstable preheater temperatures, temperature deviation was reduced by 22%, coal use by 3.2%, enabling stable control with higher alternative fuel usage.

INEEJI Inc. – Domestic Application Case

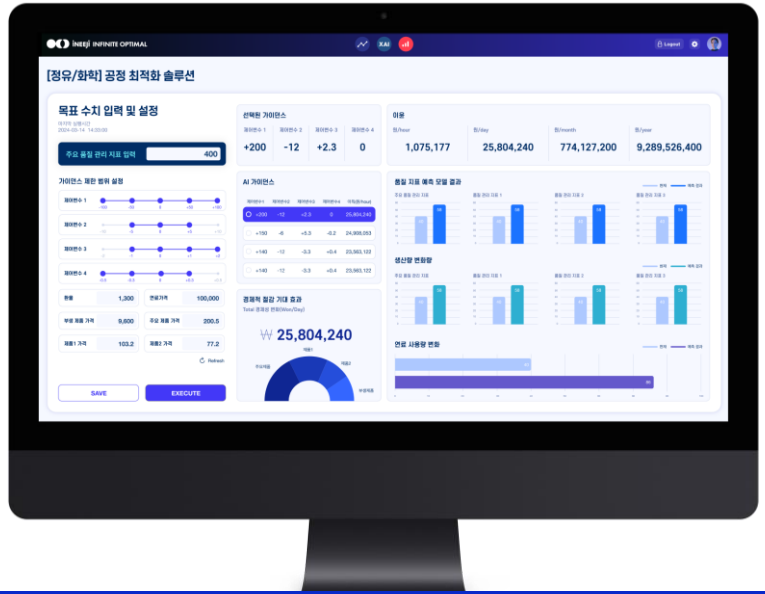
RHDS (Residue Hydrodesulfurization) Process

Diesel Quality Prediction | Product Quality Improvement & Fuel Optimization

SK Innovation Launches “Smart Plant” in Ulsan (May 24, 2024)

The company expects to cut annual costs by over 10 billion KRW. Automated process control (APC) will save around 2 billion KRW by replacing manual adjustments, while predictive maintenance solutions are projected to save an additional 2–3 billion KRW annually.

Before	Delay in quality analysis Real-time control challenging
After	75% lower prediction error Higher productivity via target quality Optimized utility costs
주요데이터	Column temperature and flow rate ※ Actual required data may vary.



INFINITE OPTIMAL SERIES™ THINK SMART

Delays in sampling and analysis hinder real-time control.
Predicting key variables (temperature, flow) improves quality, productivity, and utility efficiency.
75% lower prediction error vs. previous AI models, boosting quality competitiveness.

TOE/year **261.3**

*TOE(Ton of oil equivalent): 석유 환산톤

Entry into the Japanese market

Expanding into the Japanese Optimization Market for Sustainable Growth

철강

Strengthen Japanese sales via GS Global and local hires

김연배 일본 지사장
동경대학교 공학박사
前 과기정통부 총괄PM
前 한양대학교 AI R&D센터장
前 삼성전자 AI R&D 상무
前 NHK 주임연구원

이동철 철강 고문
한국외국어대 학사
산업포장(2022)
前 동국제강 일본/미국 법인장
前 동국제강 마케팅총괄

Sales Channel

5▲

tb innovations

Ultimatrust

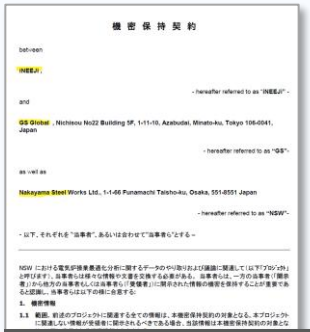
GS Global Japan

Mebius

Liberaware CO.,LTD

경제인협회
https://www.fki.or.kr › main › news › statement_detail
한경협, 日경단련과 함께 韓스타트업 일본진출 지원
2024. 4. 2. — 한경협, 日경단련과 함께 韓스타트업 일본진출 지원. - 한일 일한미래파트너십기금, 도쿄서 한일 스타트업 협력포럼 개최 -. · AI, 스마트물류, 제약, ...

뉴스1
https://www.news1.kr › articles
한국 유망 스타트업, 日대기업과 사업 협력 나선다
2024. 4. 2. — 한국경제인협회(한경협)와 일본경제단체연합회(경단련)가 공동 설립한 한일미래파트너십재단(재단)은 2일 일본 도쿄 경단련회관에서 '한일 스타트업 ...


機密保持契約

Pilot Installation

+3

23.12 OOOO 제강 시범 설치 엔화 매출 실적
24.04 OOO강철 시범설치 확정
24.05 OOO 시범설치 확정

Ciment

First Case

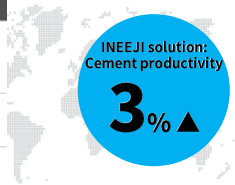
ABB digital systems enhance productivity and save energy at Tokuyama Cement processes

ABB Ability™ Expert Optimizer

Improved operations at Tokuyama Nanyo Cement, one of Japan's largest single plants, reducing kiln heat energy consumption by 3%

Expansion

구분	INEEJI	ABB
Owns Proprietary Engine	✓	✗
Provide Explainability	✓	✗
Equipment Status	✓	✓



INEEJI Inc. – Domestic Application Case

朝鮮日報

조선경제 › 산업·재계

'공정효율 스타트업' 인이지, 日 진출 1년 만에 실증 실험

산업 공정의 효율을 높이는 인공지능(AI) 기술을 제공하는 스타트업 인이지(INEEJI)는 일본 컬러 강판 제조 기업 지요다 강철 공업과 AI 공정 실증 실험을 시작했다고 6일 밝혔다.

인이지는 철강, 석유화학, 시멘트, 발전 기업 등을 대상으로 제조 효율을 높이는 기술과 클라우드 기반의 AI 설루션(Solution)을 제공하고 있다. 제철소 전기로, 유리공장 용해로, 시멘트 소성로 등 공정 변수가 복잡해 정밀한 예측·제어가 어려운 산업 현장에서 제조 데이터를 단순 활용하는 데 그치지 않고, 현장 운영자의 작업 방식과 노하우까지 AI 모델에 반영하는 것이 강점이다. 유리 공장의 경우, 용해로 내부 목표 온도를 유지하기 위한 최적의 연료 투입량을 파악하는 게 어려웠는데 AI 분석으로 연료 사용은 줄이고 생산성은 더 높이는 식이다. 인이지 관계자는 "설루션을 도입한 현장에서 품질 향상, 일관성 확보, 에너지 비용 절감 등이 입증돼 고객사가 늘고, 일본에도 진출했다"고 했다.

인이지는 작년 4월 일본 도쿄지사를 설립, GS그룹의 종합상사 GS글로벌 재팬과 협업하고 있다. 올해 상반기 지요다 강철 공업을 비롯해 시멘트 제조사, 생활가전 제조 기업 등 4사와 실증 실험을 시작한다. 지요다 강철 공업의 사카타 모토부 사장은 "인이지의 AI 기술을 활용해 품질 향상을 극대화하고, 고객사의 친환경 요구에도 효과적으로 대처해 나가겠다"고 했다. 최재식 인이지 대표는 "다양한 업종의 생산 공정에서 탄소 배출을 줄이고, 에너지 저감 효과를 인정받았다"며 "일본 진출은 선진국 시장 개척을 위한 교두보 역할을 할 것"이라고 했다.

- Japan has **high entry barriers, requiring localization and trust**.
- Japanese companies and consumers value quality and reliability; **quick contract wins show domestic technology meets their standards**.
- As of 2024, INEEJI leads in domestic AI-based autonomous manufacturing pilot clients.
- Starting with Japan, it aims to expand its global network.