

# Building a Common-Ground for Mutual Understanding Between Humans and AI

Yasuhiro Katagiri

Artificial Intelligence Research Center  
AIST, Japan

# LLM as a partner?


# Vanilla LLM doesn't care who/what you are talking about

- **User:** Can you suggest a **travel plan to Taiwan for three days**?  
Request for a travel plan suggestion
- **LLM:** Absolutely! Here's a three-day travel plan for Taiwan:  
LLM responded with a reasonable plan
- ... <Travel itinerary suggestion> ...
- **User:** What about the effects of **the earthquake**?  
I was asking about a specific earthquake
- **LLM:** Taiwan is located in a seismically active region,  
LLM mistook my intent and start giving a lot of generic and irrelevant information
- ... <General description of effects of earthquake on travel in Taiwan> ...
- **User:** What itinerary are you talking about?
- **LLM:** My apologies for the confusion.
- ... <Excuse and additional advice information> ...  
I repeated the request for the same specific earthquake
- **User:** I am concerned about **the earthquake that occurred in Hualien**.  
LLM misidentified the earthquake I was referring to and started giving a lot of irrelevant information
- **LLM:** I understand your concern. In February 2018, a magnitude 6.4 earthquake struck near Hualien City in eastern Taiwan.
- ... <Explanation of an old earthquake and general advice on travel to Hualien after an earthquake> ...  
I repeated again the request for the same specific earthquake
- **User:** I'm talking about **the very recent one**. Do you have any information on the damage it caused?
- **LLM:** I'm sorry for the oversight.  
LLM gave up
- ... <Excuse and further advice information> ...

# Current LLM as a partner

- LLM is highly knowledgeable and can give us potentially useful information.
  - One way push of abundant and supposedly helpful information
  - Users burdened to make sense of voluminous LLM output
  - Skills needed to provide appropriate queries/prompts
- LLM gives uncooperative or patronizing impression to users.
- LLM does not care about building and utilizing common ground.

# Basic common ground maintenance

- (1) a. A: Did you see Sato-san in the meeting yesterday?  
b. B: Yes, he seemed to be happy because of his recent promotion.  
c. A: He got promoted?  
d. B: Didn't you mean Yo Sato?  Misidentification of "Sato" could stay here  
e. A: Um, I meant Toshi Sato.  
f. B: Oh, sorry. No, I didn't see him yesterday.

Common ground presupposes higher-order cooperation

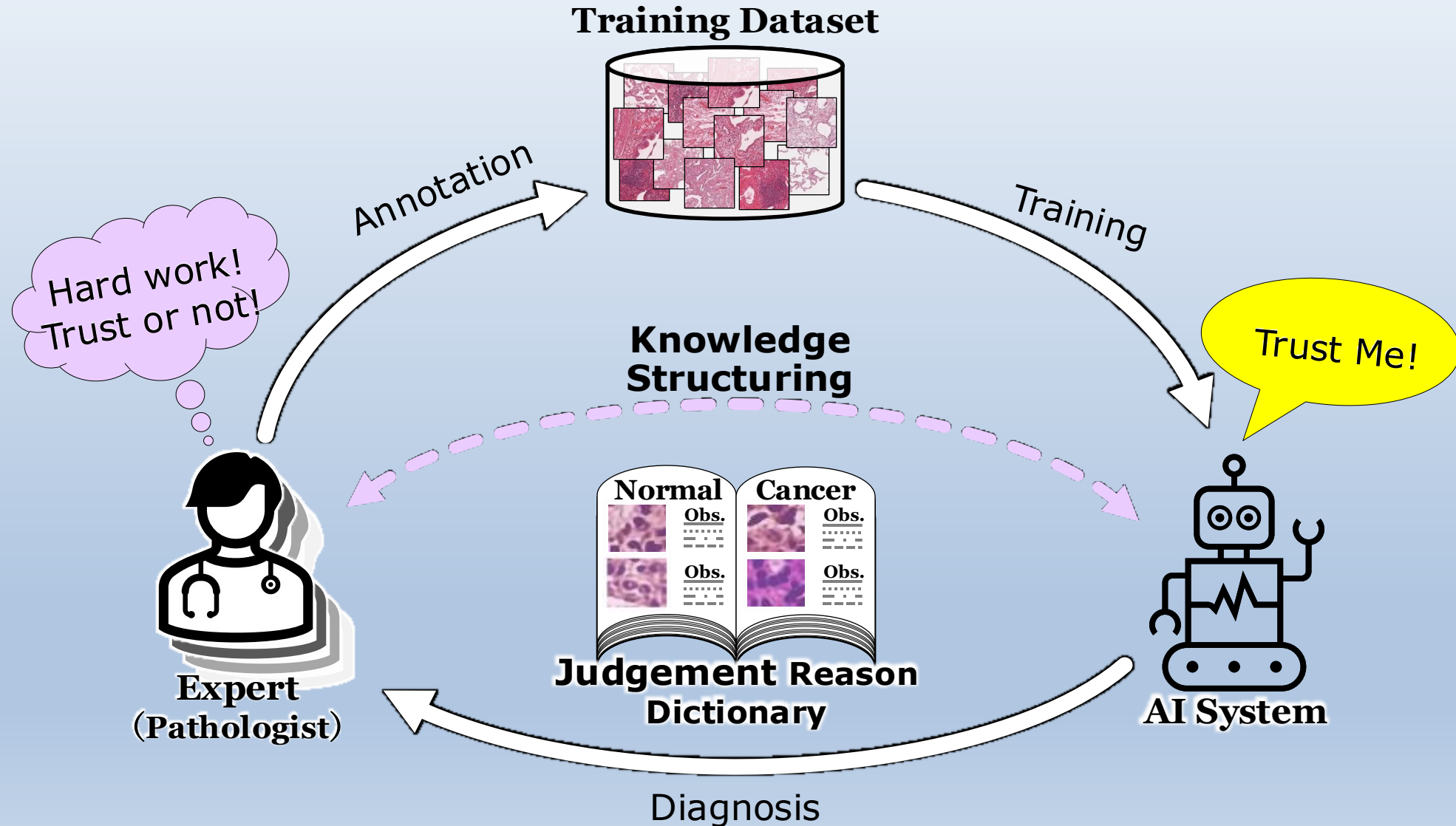
A and B are both responsible for maintaining grounding status (referring to the same person)

# Characteristics of Human collaboration

- Higher-order cooperation
  - Participants are committed to and responsible for collaborative endeavor
- Shared commitment for task completion
  - Shared goal
  - Plan breakdown and task allocation
  - Coordinated execution
- Constant negotiation
  - Information alignment
  - Intention alignment

# Toward Human-AI common ground: Judgment reason dictionary for medical image diagnosis

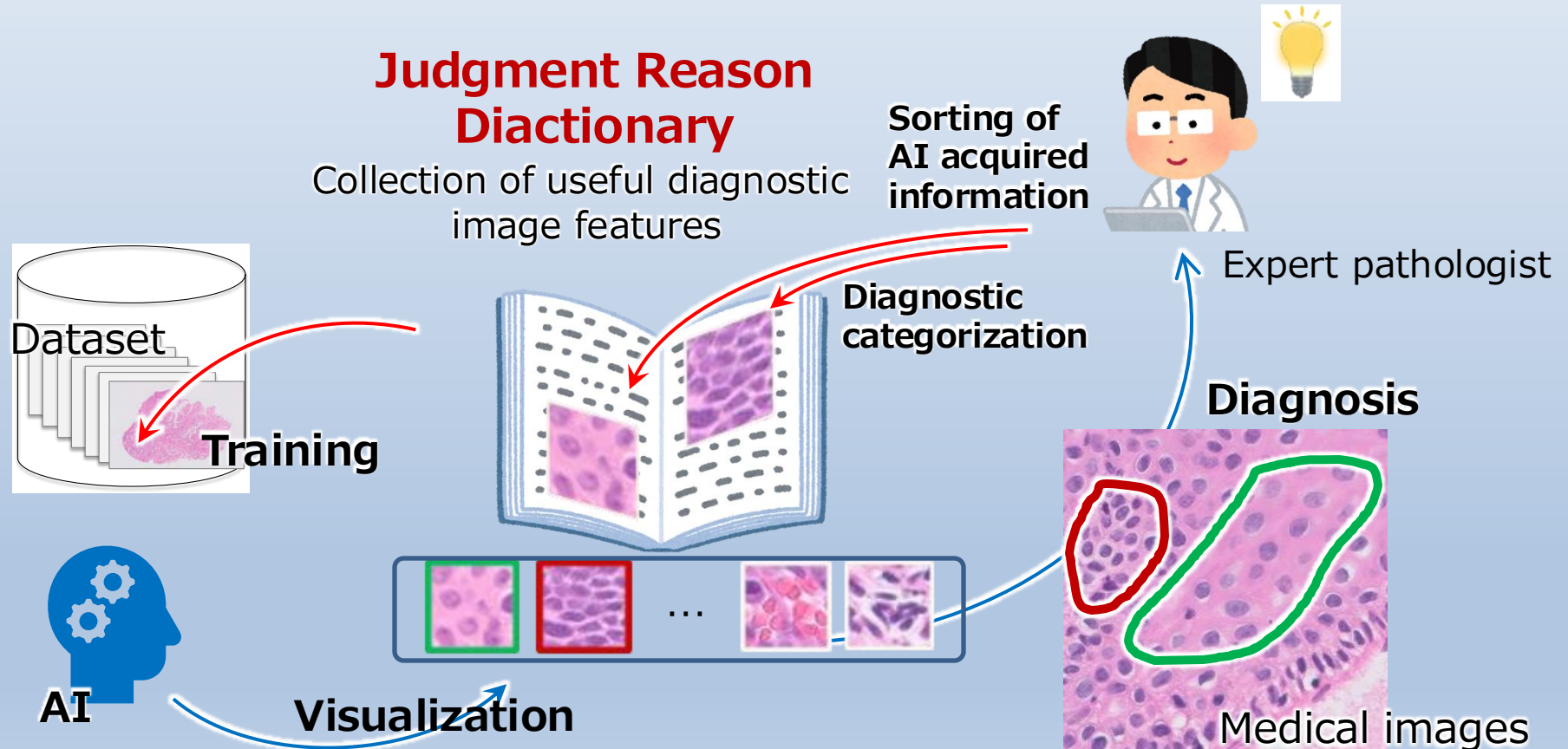
# Medical Image Diagnosis



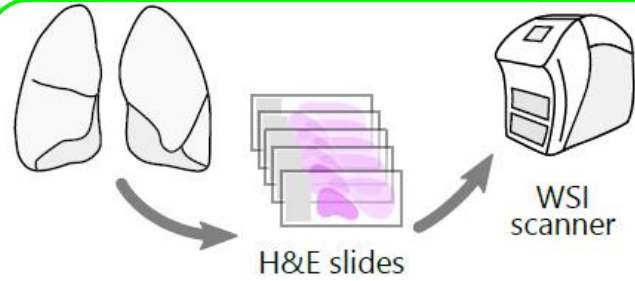


# A Basis for Human-AI collaboration

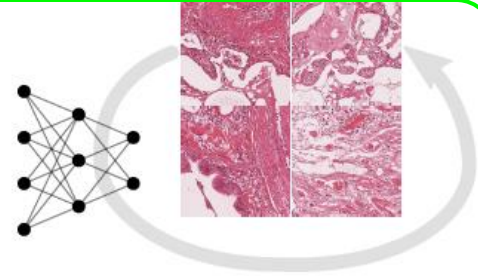
Visualize AI acquired information and sort them by human expertise



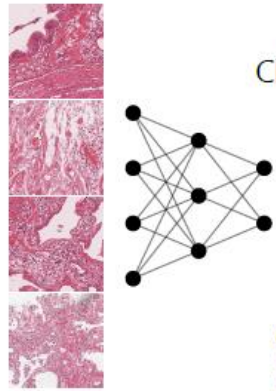
### 1. Image Preparation



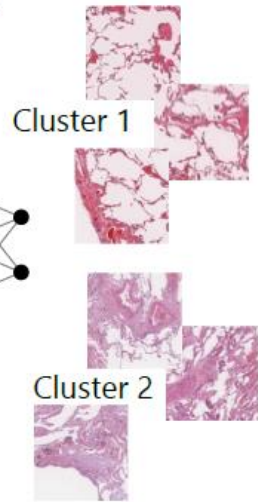
### 2. Self-supervised Learning



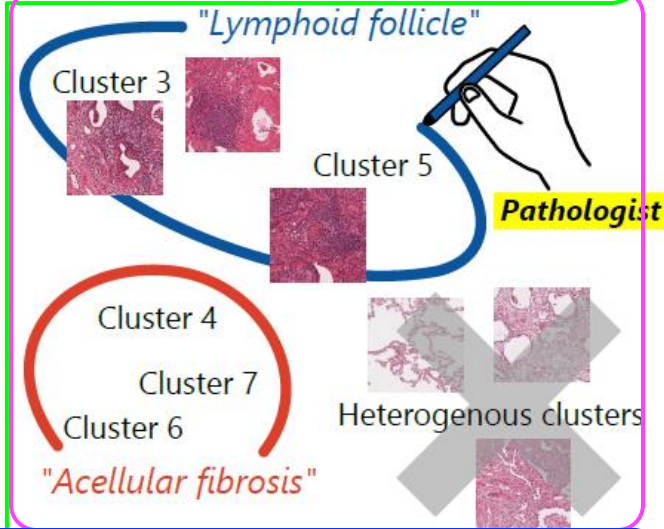
### 3. Feature Extraction



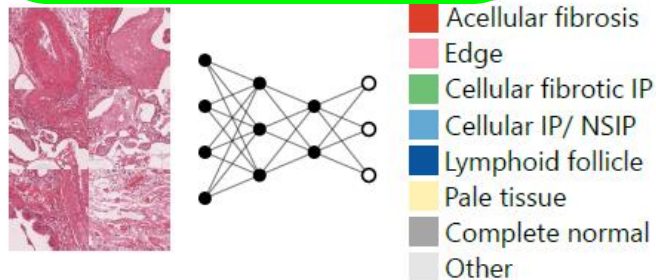
### 4. Clustering



### 5. Cluster Integration



### 6. Transfer Learning

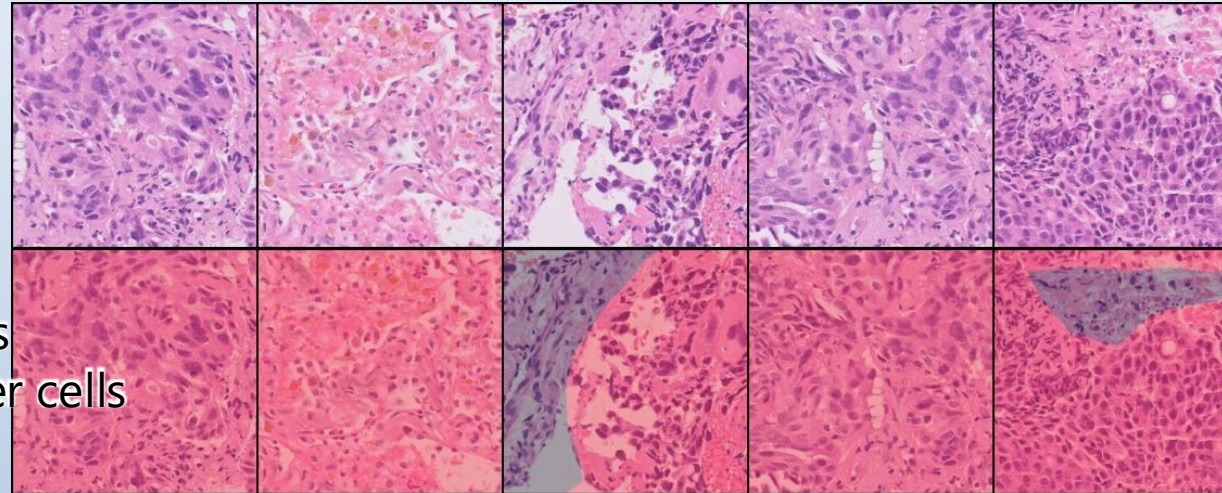


### 7. Mapping

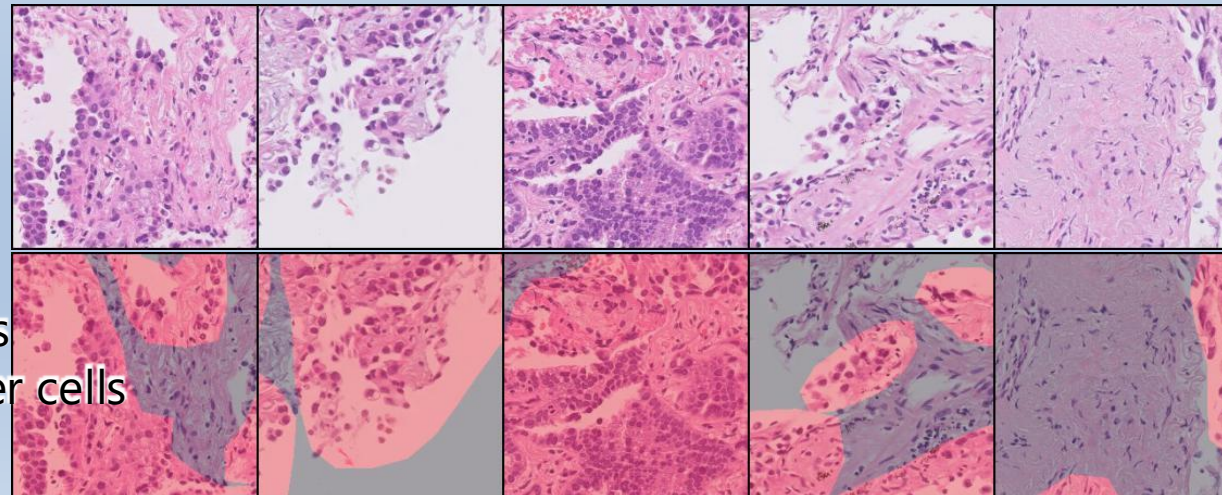


# Cancer positive dictionary items

red region shows  
location of cancer cells

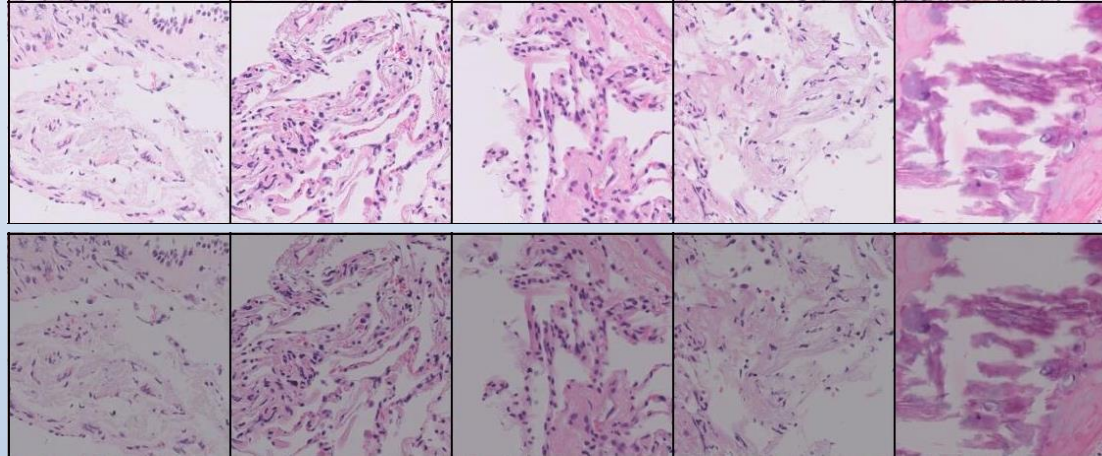


red region shows  
location of cancer cells

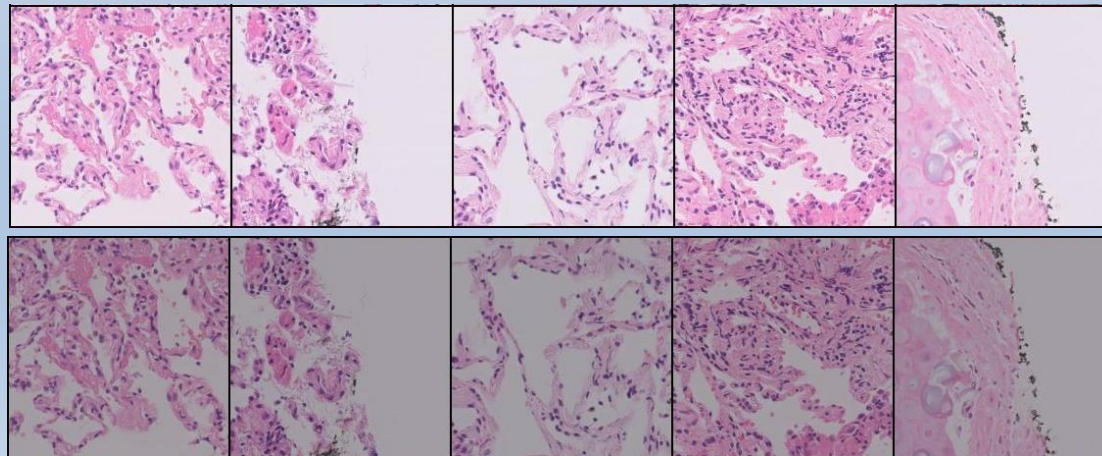


# Cancer negative dictionary items

no cancer cells



no cancer cells



# Image diagnosis flow

## Explainable feature representation:

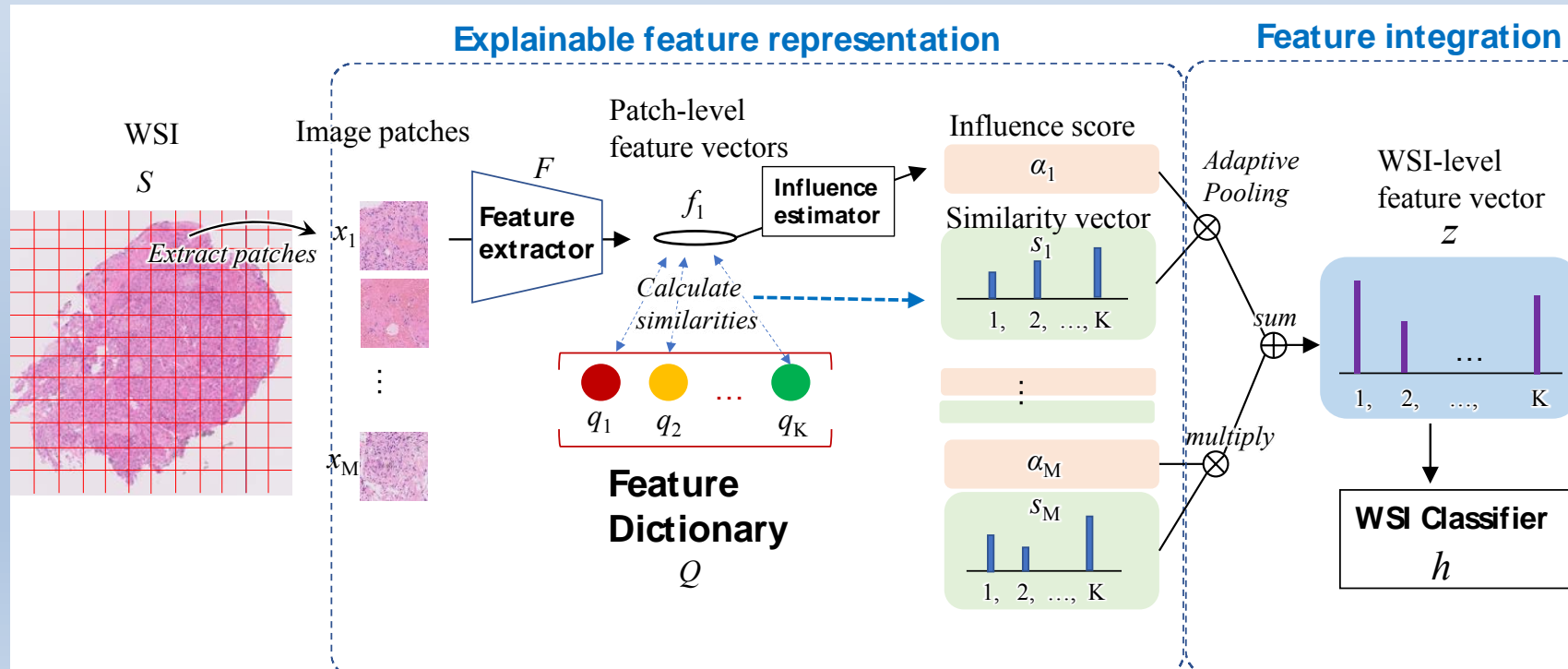
Local region image patches are represented by dictionary entry similarities

## Feature integration

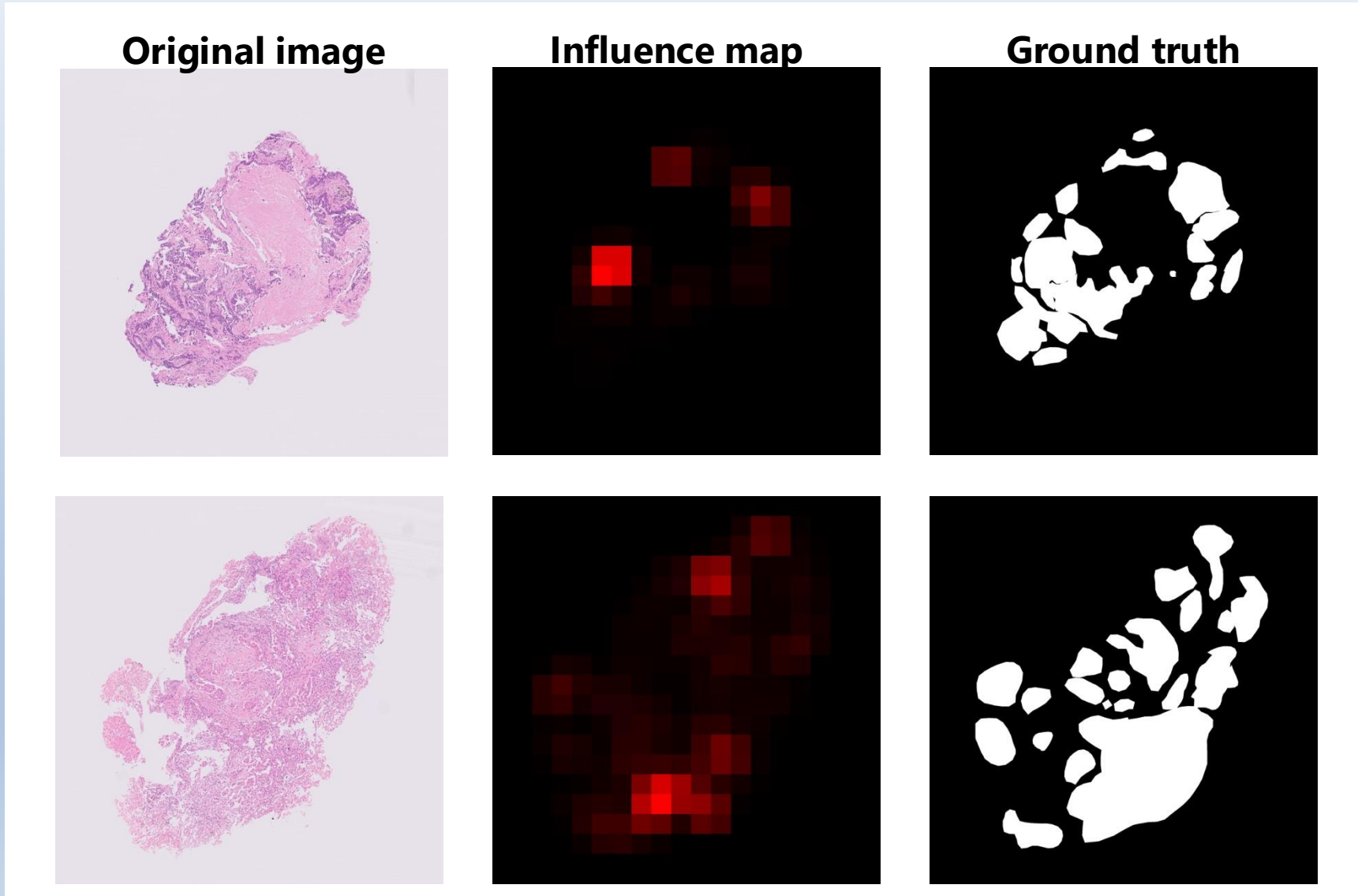
Representation of local regions are combined according to their significance

## Whole image classification

Combined feature vectors are fed into a classifier



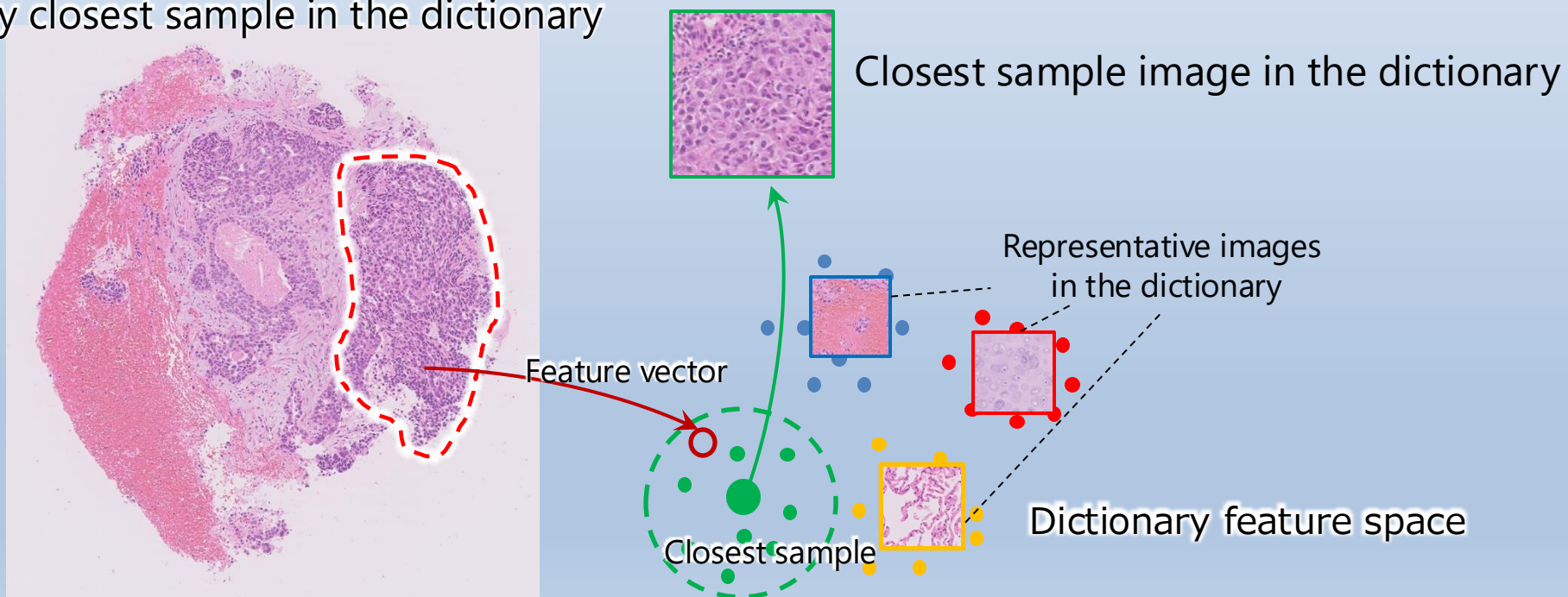
# Estimated region of significance



# Whole slide image diagnosis with judgment reason

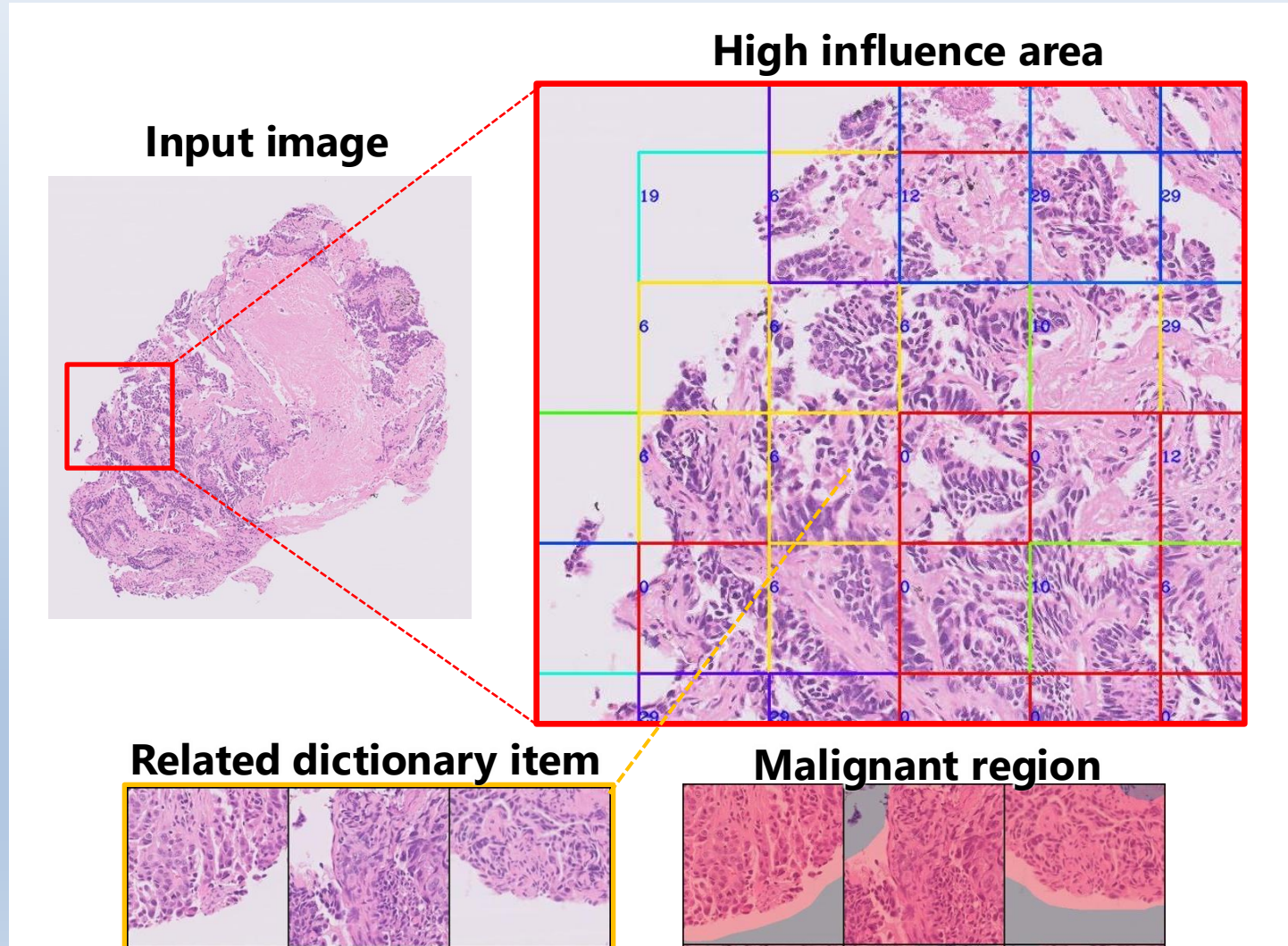
- Judgment reason dictionary consists of representative images of target pathologies
- Provide samples from judgment reason dictionary that contributed to the diagnosis
- Users (medical doctors) can understand the reason behind AI judgments

Identify significant region of influence and identify closest sample in the dictionary



# Whole slide image interpretation

Dictionary entry closest to the significant region includes cancer cell images





# Summary

- Building and maintaining Common Ground is essential for transparent AI.
- Engaging in a joint task with AI, medical image diagnosis, conversations or other tasks, will greatly be facilitated by common grounding, e.g., to disclose their reasoning and preferences and try to align their behaviors toward their shared goals.
- Common ground presupposes higher-order cooperation between participants, both humans and AI.
- Common grounding in AI will also be essential for establishing trust in AI.

Thank You