PRE-UNIVERSITY SUMMER SCHOOL · SEMINAR

AI Seminar

Language models, generative vision models

AI

Paris Giampouras · Pre-University Summer School

Can a machine create something genuinely new?

What ‘new’ really means?

We will learn by predicting, testing and improving

01 Language: next-token prediction

02 Vision: generating from a learned space

03 Use case: image classification

04 Use case: recommendation systems

Predict the next word

Complete this sentence:

"The student opened the…"

Why did several different answers all feel reasonable?

What information would change your prediction?

Language models use tokens

VISIBLE TEXT

unbelievable!

Humans see a word, meaning and punctuation.

TOKENS ARE CHUNKS OF WORDS

un · believ · able · !

A tokenizer maps pieces to integer IDs. The model processes those IDs as vectors.

A language model assigns probability—not certainty

The student opened the

Possible next tokens

Bigram Model: A simple language model

Bigram models count word pairs:

P(next word | previous word) = pair count ÷ previous-word count

Bigram model: is tiny and limited—but it exposes the same generate-one-token-and-repeat loop used at much larger scale.

A bigram model predicts from one token of context

01

A bigram is a pair of neighbouring tokens

AI | can | help

bigram 1: AI can

bigram 2: can help

A bigram model predicts from one token of context

01

A bigram is a pair of neighbouring tokens

AI	can	help

bigram 1: AI can

bigram 2: can help

02

Count what appears after “can” in our corpus

AI can help
Students can help
AI can create
AI can learn

PAIR COUNTS

can → help	2	**can → create**	1
		can → learn	1

A bigram model predicts from one token of context

01

A bigram is a pair of neighbouring tokens

AI can help

bigram 1: AI can

bigram 2: can help

02

Count what appears after “can” in our corpus

AI can help
Students can help
AI can create
AI can learn

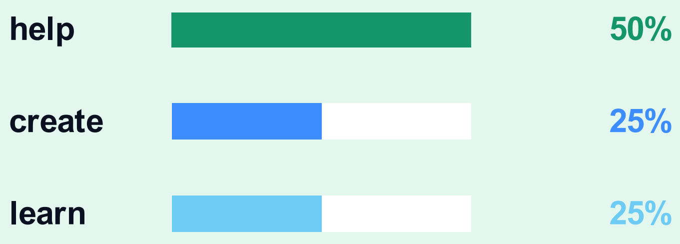
PAIR COUNTS

can → help	2
can → create	1
can → learn	1

03

Convert counts into next-word probabilities

$P(\text{help} \mid \text{can}) = 2 \div 4 = 50\%$



GENERATE sample one word → append it → use the new last word → repeat

LIMIT only one token of context

Temperature Parameter in Language Models

Temperature reshapes the probabilities before the model samples its next token.

T = 0.25

FOCUS

Sharpened distribution

help	85%
create	5%
build	5%
ask	5%

More repeatable · top choice dominates

T = 0.8

BALANCE

Moderately concentrated

help	43%
create	19%
build	19%
ask	19%

Some variety · still follows strong preferences

T = 1.5

EXPLORE

Flatter distribution

help	34%
create	22%
build	22%
ask	22%

More variety · unlikely choices appear more often

LOWER T ← focus

T = 1 uses the model's original distribution

explore → HIGHER T

Temperature changes diversity—not knowledge

LOW TEMPERATURE

Concentrated choices

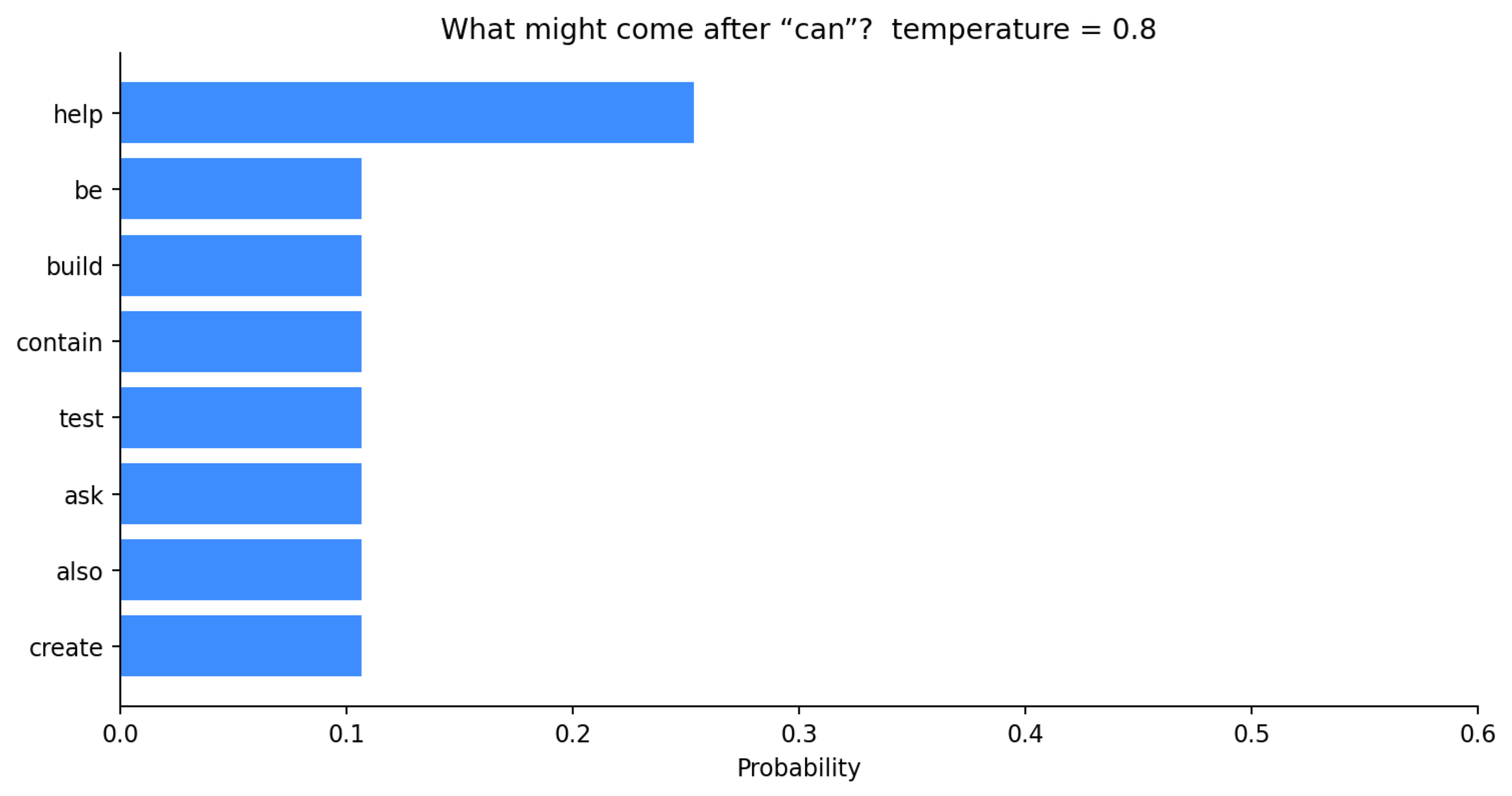
High-probability tokens dominate. Output is more repeatable but can become dull or stuck.

HIGH TEMPERATURE

Flatter choices

Less likely tokens appear more often. Output is more varied, but errors and incoherence become more likely.

Demo · Watch next-token probabilities become a sentence



The role of temperature parameter

Different choices for different context:

- a factual definition
- five story ideas
- computer code
- a fictional character name

Why does creativity sometimes benefit from uncertainty while factual work usually does not?

From Bigrams to Neural Language Models

BIGRAM

Looks one token back

Fast and understandable, but loses topic, grammar and long-range relationships.

From Bigrams to Neural Language Models

BIGRAM

Looks one token back

Fast and understandable, but loses topic, grammar and long-range relationships.

NEURAL LM

Builds contextual representations

Many layers combine information from multiple tokens to score what should come next.

Embeddings

Embeddings turn discrete tokens into coordinates

token ID → vector of learned numbers (embedding)

Tokens used in similar contexts often develop related representations.
Contextual models then update those representations based on the surrounding sequence.

Attention Mechanism and Transformers

Attention mechanism

Let the following sentence:

“The robot lifted the suitcase because it was blocking the door.”

When ‘it’ appears, which earlier words matter most?

Attention lets each token gather relevant context

Query (Q)

What information does this position need?

Keys (V)

What information can other positions offer?

Weights

How relevant is each position?

Mix

Combine values (V) using those weights

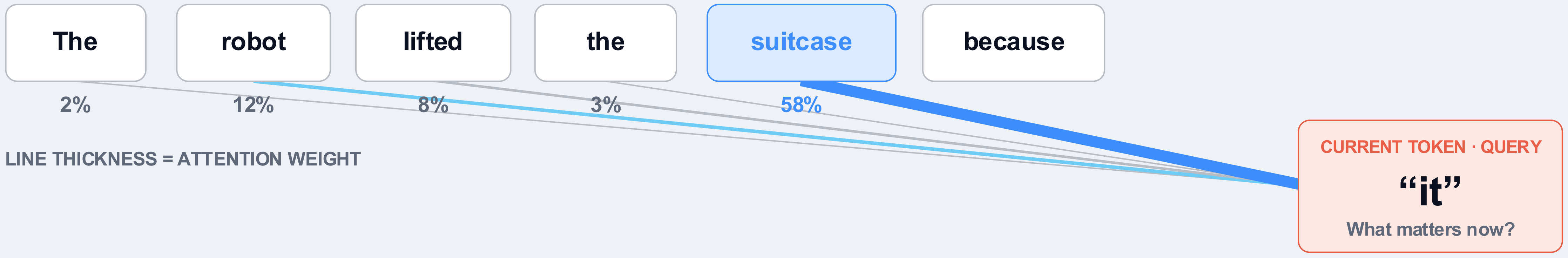
Different attention heads can learn different relationships.

Attention mechanism

To update “it”, one attention head compares its query with keys from itself and earlier tokens, then mixes their values.

ONE ILLUSTRATIVE ATTENTION HEAD

EARLIER TOKENS · KEYS + VALUES



QUERY

What information does ‘it’ need?

KEYS

What can each earlier token offer?

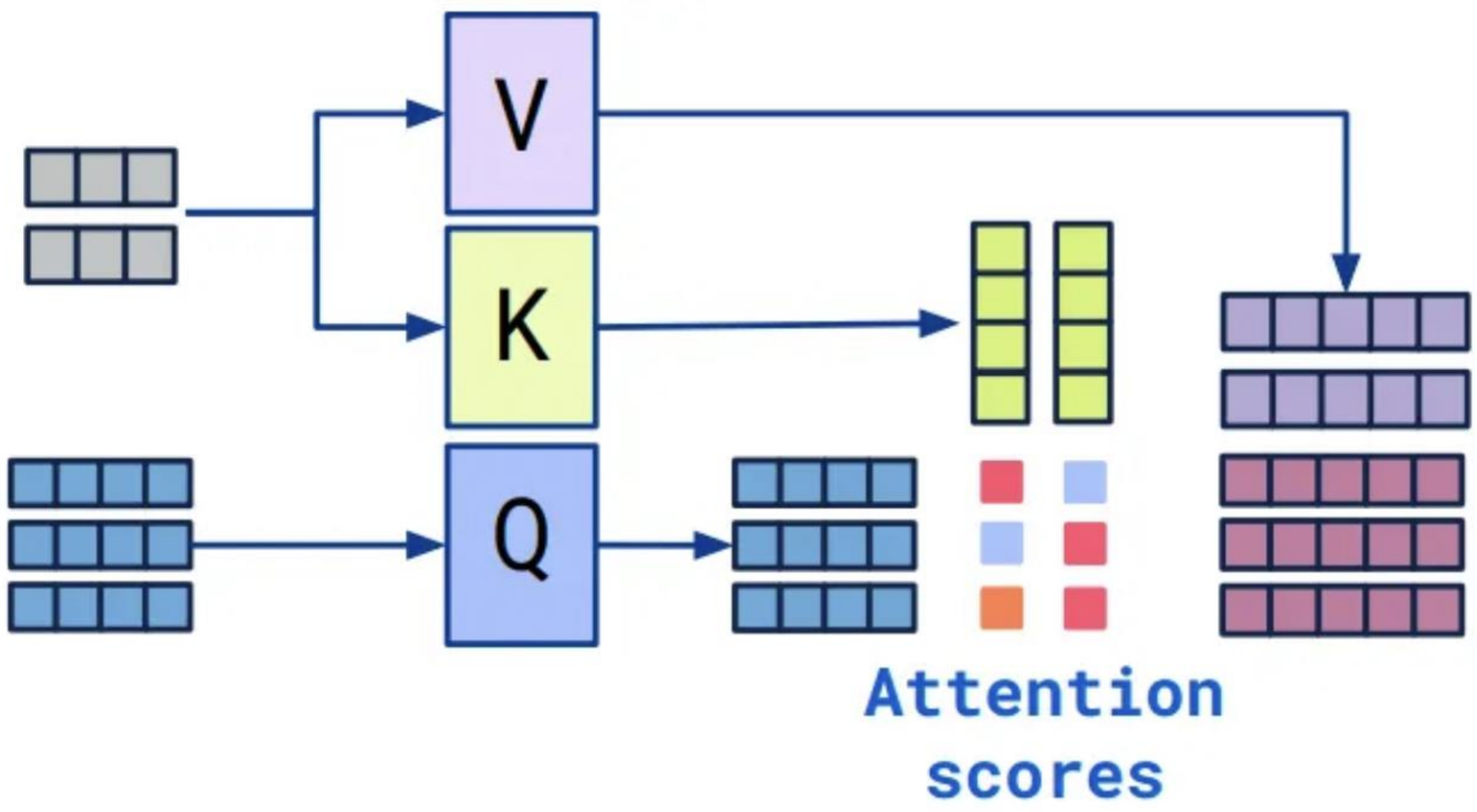
WEIGHTS

Stronger similarity means more influence.

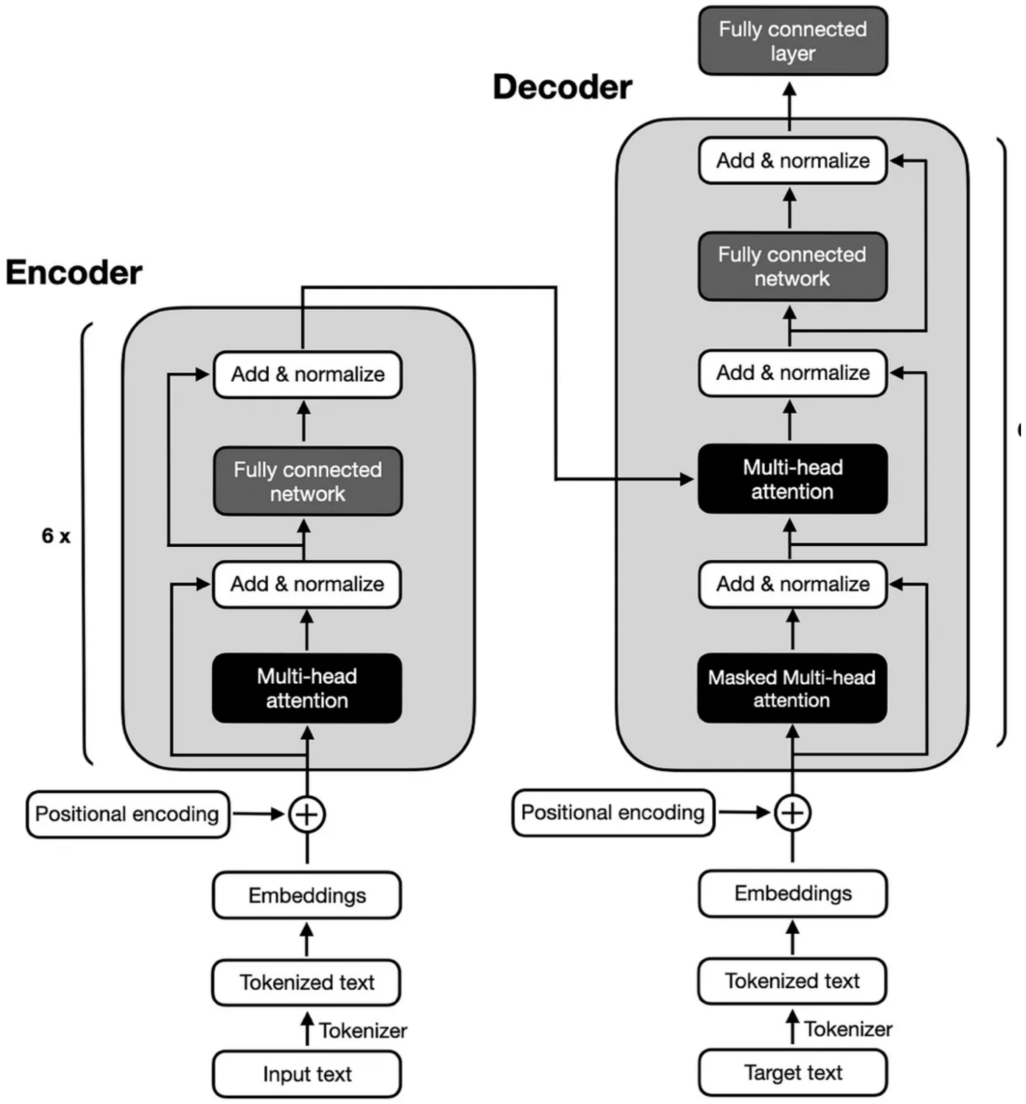
VALUES

Mix token information using those weights.

Attention lets each token gather relevant context



Transformers



Transformers repeatedly update token representations

01

Embed

Tokens become vectors plus position information

02

Attend

Share relevant context across positions

03

Transform

Apply a learned non-linear update

04

Predict

Convert the final state to token scores

Residual connections and normalisation help deep stacks train reliably.

Pretraining and Post-training

PRETRAINING

Predict tokens across large datasets

Learns language, knowledge-like associations, code patterns and general representations from varied sequences.

Pretraining and Post-training

PRETRAINING

Predict tokens across large datasets

Learns language, knowledge-like associations, code patterns and general representations from varied sequences.

POST-TRAINING

Practice useful responses

Reasoning, Instruction examples, preference feedback, safety tuning and task-specific adaptation.

Prompting

How to write the right prompt?

Task	Context	Constraints	Format
State exactly what should be done	Provide relevant facts and audience	Define limits, tone and exclusions	Specify the desired structure

A strong prompt reduces ambiguity

Example of a vague prompt

Improve this request:

“Explain climate change.”

Add an audience, purpose, scope, evidence requirement and output format.

Which addition changes the answer most? Which part helps you verify quality?

A good prompt is specific and checkable

The topic is unchanged. The added details reduce guesswork and create visible quality criteria.

IMPROVED PROMPT

AUDIENCE

Explain climate change to a 15-year-old who knows basic science.

PURPOSE + SCOPE

Help them prepare for a classroom discussion. Cover the greenhouse effect, why scientists attribute recent warming mainly to human activity, two likely impacts by 2050, and one common misconception.

EVIDENCE

Separate well-established evidence from uncertainty. Use three reliable, recent sources and provide links.

FORMAT

Write 250–300 words in plain language. Define technical terms and finish with four takeaway bullets.

WHAT CHANGED?

BEFORE "Explain climate change."

- ✓ **Audience** — 15-year-old learner
- ✓ **Purpose** — classroom discussion
- ✓ **Scope** — four required ideas
- ✓ **Evidence** — sources + uncertainty
- ✓ **Format** — length + final bullets

Dealing with hallucinations

WHEN GROUNDED

Evidence constrains the answer

The model works from supplied documents, retrieved sources, tool outputs or verified data.

WHEN UNGROUNDED

Plausibility can replace evidence

The model may invent a citation, merge facts or confidently answer an ambiguous question.

Grounding: Putting external evidence in context

A person can supply the source, or a tool used from the AI model can retrieve it.

SAME QUESTION **"According to the school report, why did attendance fall in March?"**

GROUNDED

External evidence enters the context

USER OR SYSTEM
uploads, pastes or attaches a source

RETRIEVAL TOOL
finds material in a trusted source

EVIDENCE NOW IN THE MODEL'S CONTEXT
"March attendance fell after two snow-closure days and a flu outbreak."

AI

ANSWER BASED ON THAT EVIDENCE
"The report attributes the fall to snow closures and a flu outbreak."

✓ We can compare the claim with the source.

UNGROUNDED

No external evidence enters the context

NO SUPPLIED SOURCE
nothing relevant was attached or pasted

NO RETRIEVAL
no tool searched for supporting evidence

EVIDENCE IN CONTEXT
"Only the question and learned patterns are available."

AI

PLAUSIBLE-SOUNDING GUESS
"Attendance probably fell because students were tired after exams."

! There is no evidence trail for the claim.

Grounded does not mean automatically true—it means the claim can be traced to evidence and checked.

Grounding connects generation to external evidence

Question

Clarify what must be answered

Retrieve

Find relevant trusted material

Generate

Use the evidence in context

Verify

Check citations and unresolved uncertainty

Retrieval reduces some errors, but irrelevant or poor sources can still mislead.

We will learn by predicting, testing and improving

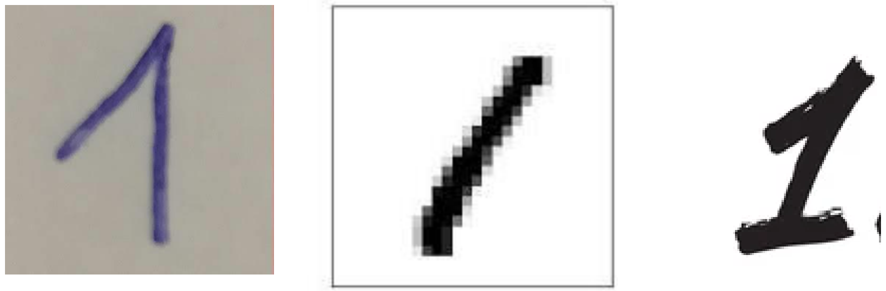
01 **Language: next-token prediction**

02 **Vision: generating from a learned space**

03 **Use case: image classification**

04 **Use case: recommendation systems**

What makes an image look like a handwritten digit?



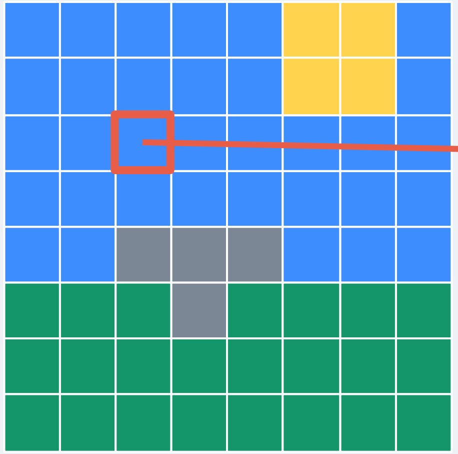
Shape? Stroke thickness? Position? Contrast?
Which details can vary while identity remains?

A generative model learns what can vary, and what must remain!

Images are grids of numbers

In an RGB image, each pixel contains three values describing how much red, green and blue it contains.

1 · IMAGE



8 rows × 8 columns

illustrative pixel art

2 · ONE PIXEL

pixel at row 3, column 3

RED	GREEN	BLUE
61	141	255

3 · THREE COLOUR CHANNELS

RED CHANNEL
61 out of 255

GREEN CHANNEL
141 out of 255

BLUE CHANNEL
255 out of 255

256 rows × 256 columns × 3 channels = 196,608 numbers

From Pixel to Latent Space

PIXEL SPACE

Many detailed coordinates

Every pixel is explicit. Small shifts can change many numbers even when the object remains the same.

LATENT SPACE

Fewer learned coordinates

Directions can capture broader factors such as thickness, pose, colour, style or other semantic features.

Autoencoders: Learning to compress and reconstruct

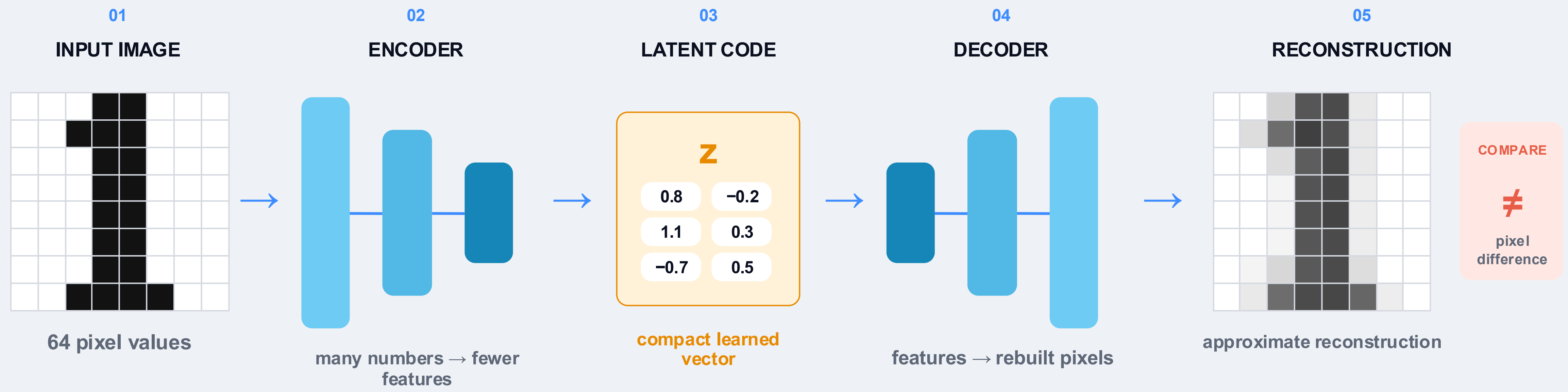
Image → **Encode** → **Decode** → **Compare**

- **Image**: Start with pixels
- **Encode**: Map to a smaller latent vector
- **Decode**: Rebuild the image
- **Compare**: Train to reduce reconstruction error

A useful latent space keeps important structure while discarding some detail.

An autoencoder learns to compress and reconstruct

It learns a compact latent representation by trying to reproduce its own input.



TRAINING LOOP

reconstruction error = difference(input, reconstruction)

adjust encoder + decoder → repeat

The bottleneck forces the model to keep useful structure instead of copying every pixel.

Diffusion Models

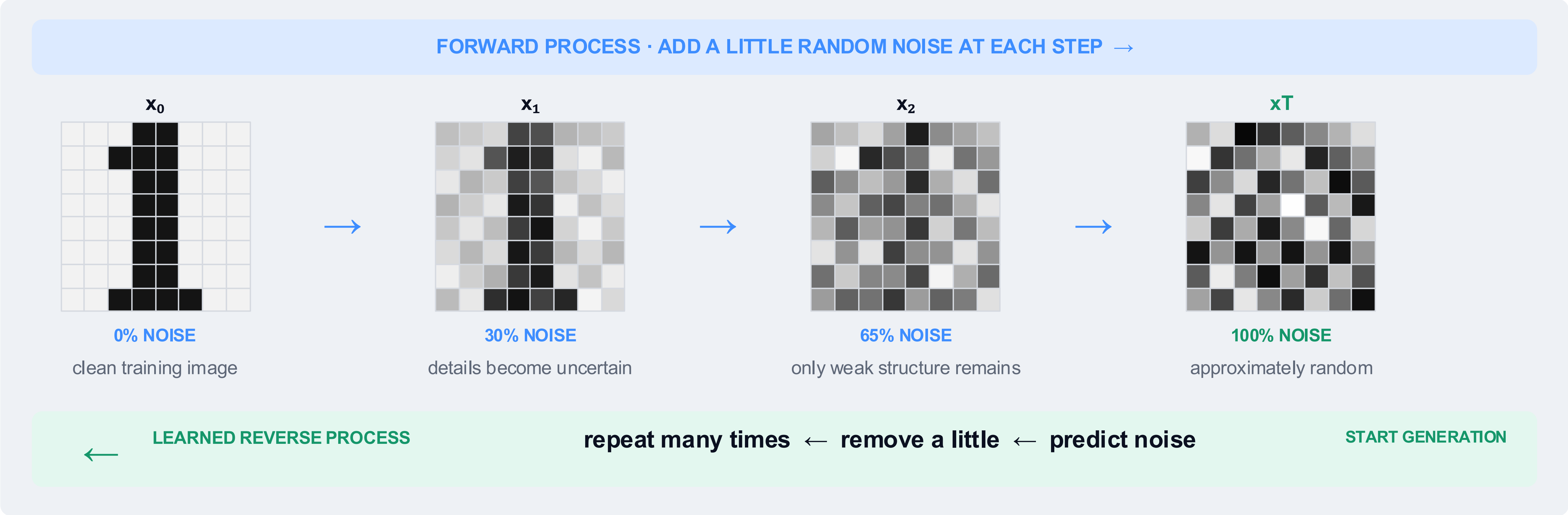
Forward process

0% noise	25%	60%	100%
Original structure is clear	Details become uncertain	Only weak structure remains	Approximately random noise

The forward process is easy; learning the reverse process is the challenge.

Diffusion learns to reverse a gradual noising process

Generation starts from noise and removes it step by step.



WHY TRAINING WORKS

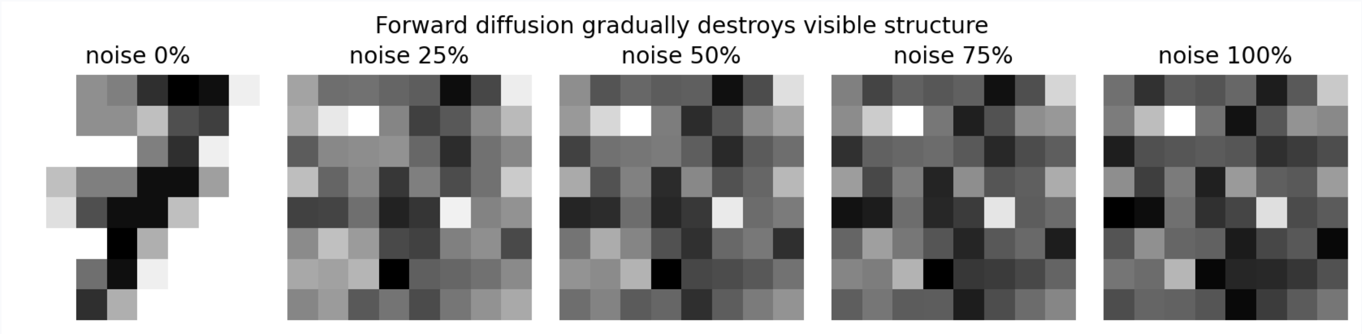
The added noise is known, so the model can practise predicting it.

WHAT GENERATION DOES

It repeats a learned denoising step; it does not jump from noise to image once.

Forward diffusion gradually hides information

DEMO



Can you still identify the digit?

Generation follows the learned reverse direction

Sample noise → **Predict** → **Refine** → **Repeat**

- **Sample noise**: Choose a random starting point
- **Predict**: Estimate the noise or clean direction
- **Refine**: Take one denoising step
- **Repeat**: Structure gradually appears

How new images are generated?
Different random seeds can produce different valid images from the same prompt.

Guidance/Steering of the Generative Process

CONDITION

Prompt or other signal

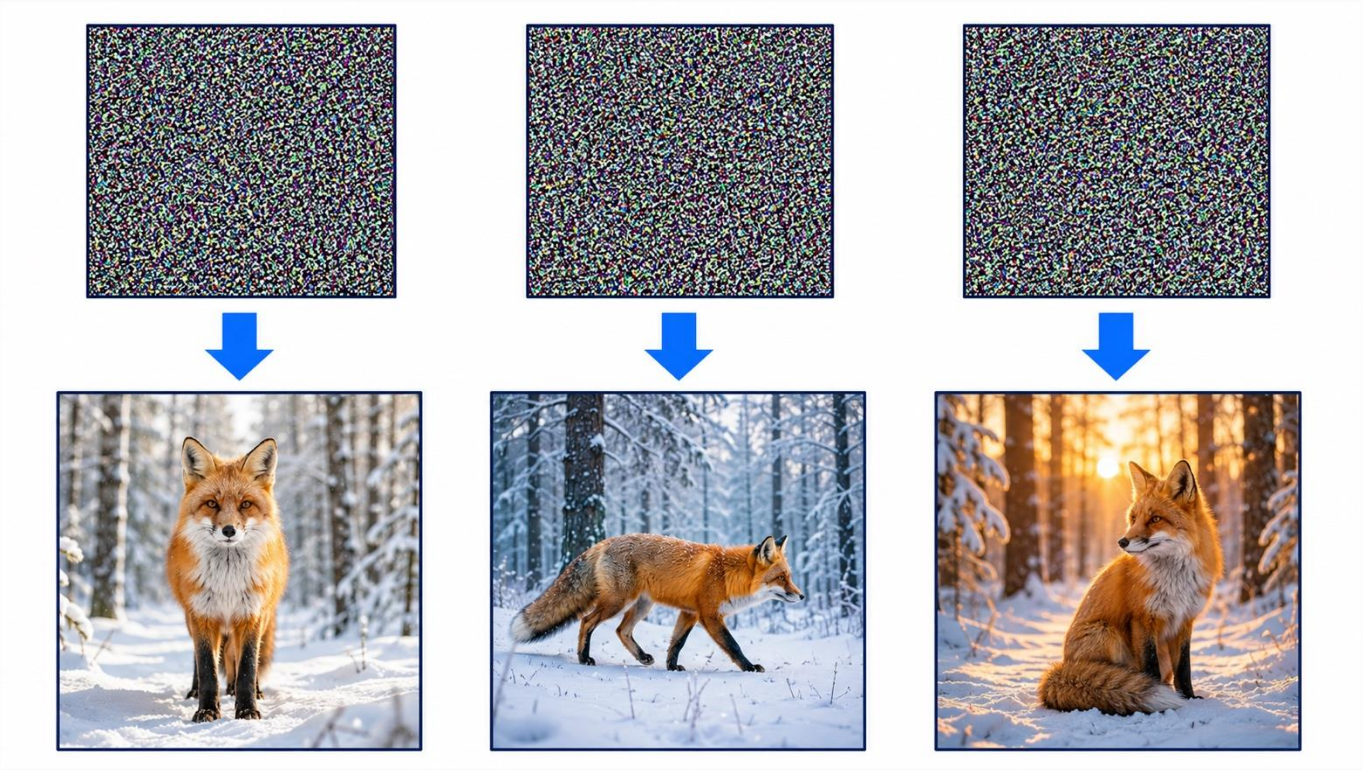
A text embedding, class label, sketch, depth map or reference image which indicates desired content.

SAMPLING

Noise provides variation

The model follows a probabilistic path, balancing prompt alignment with learned visual plausibility.

Same prompt, different seeds lead to different valid images



ONE CONDITION · MANY POSSIBLE SAMPLES

“A red fox in a snowy forest”

01 CONDITION

Keeps the subject and setting aligned.

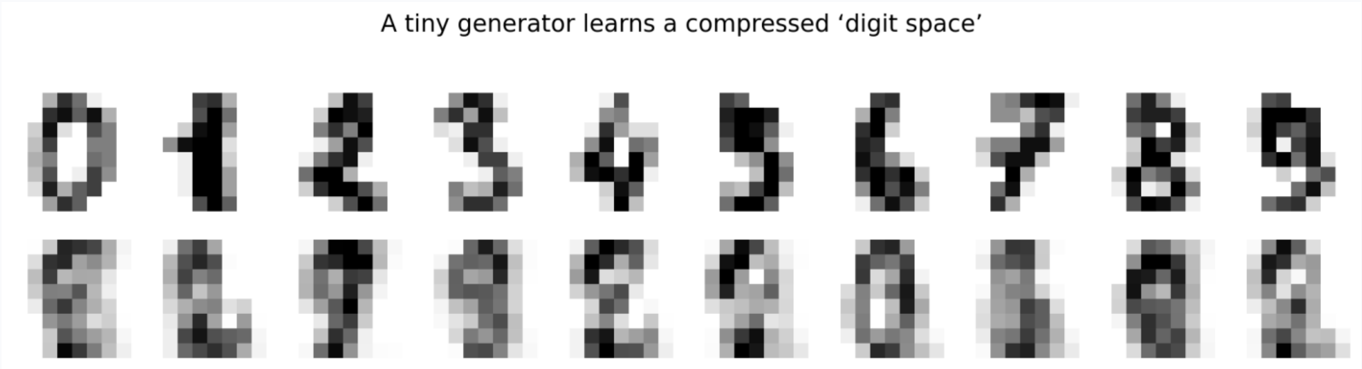
02 SEED

Changes the starting noise and the denoising path.

03 RESULT

Pose, composition and lighting can differ.

Demo 2 · Sample a tiny learned image space



DEMO

Real digits above; generated samples below

Image generation creates new risks as well as new tools

OPPORTUNITY

Rapid visual exploration

Storyboards, concept art, accessibility, design variations and synthetic training examples.

RESPONSIBILITY

Provenance and consent

Misleading media, impersonation, biased representation, creator rights and unclear disclosure require safeguards.

We will learn by predicting, testing and improving

01 **Language: next-token prediction**

02 **Vision: generating from a learned space**

03 **Use case: image classification**

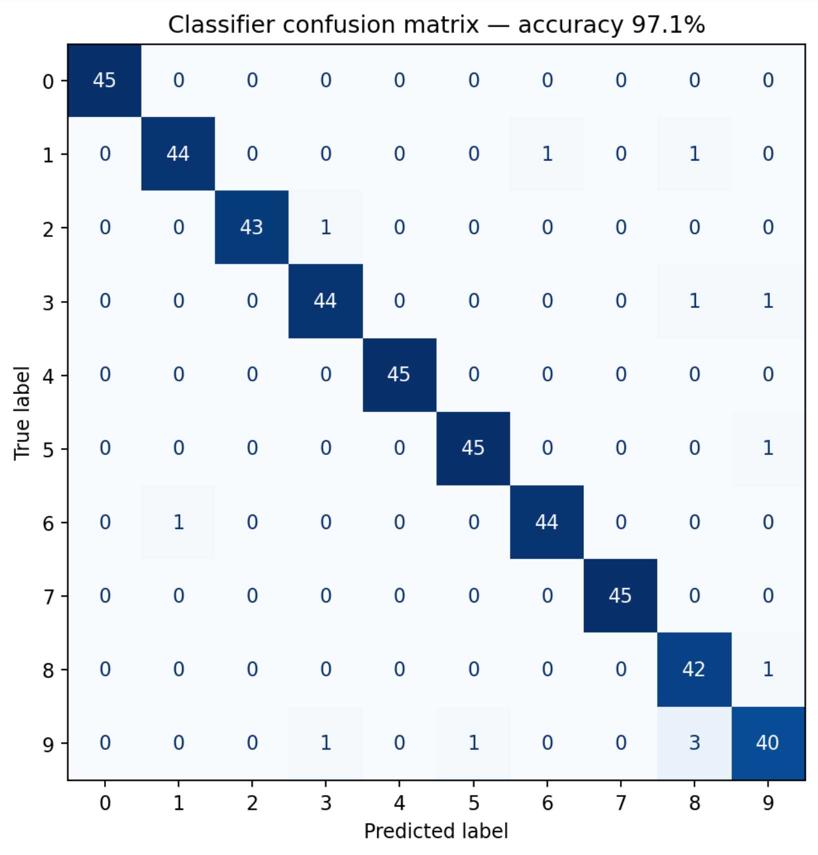
04 **Use case: recommendation systems**

Classification asks a different question from generation

Given this image, which label is most plausible?

The output is a distribution over classes. Accuracy summarises performance.

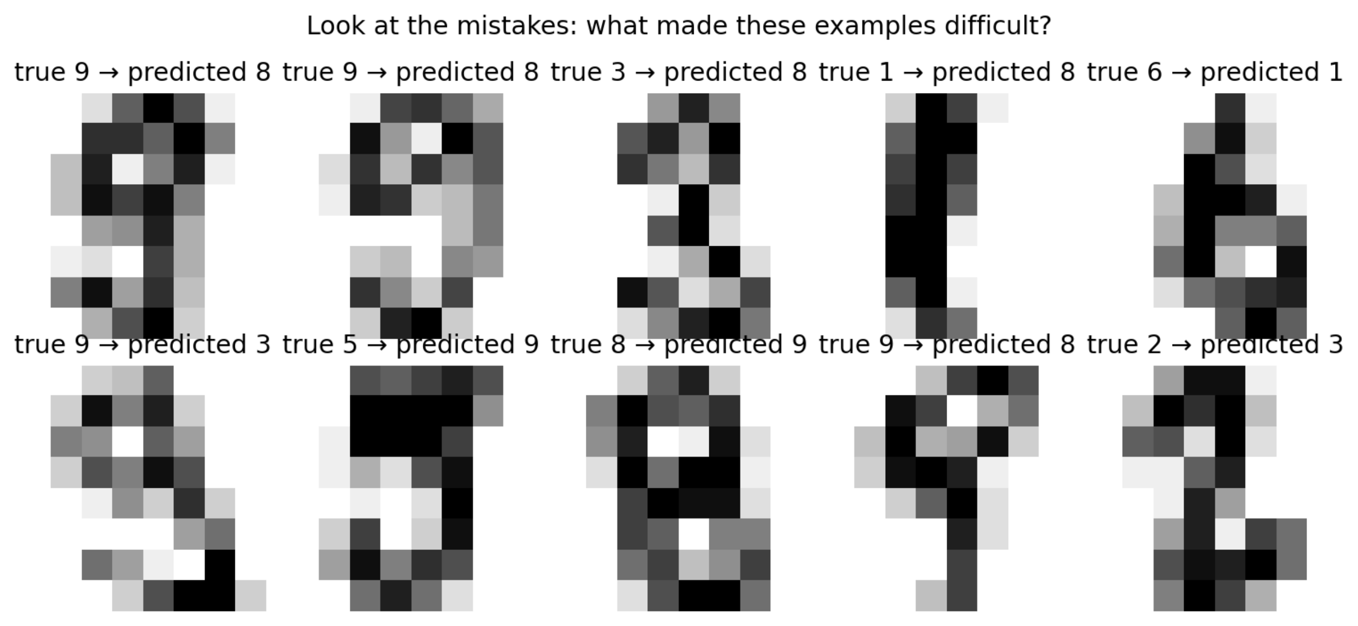
Demo 3 · A digit classifier is accurate—and still wrong sometimes



DEMO

Confusion matrix

Mistakes are clues about the model and the data



We will learn by predicting, testing and improving

01 **Language: next-token prediction**

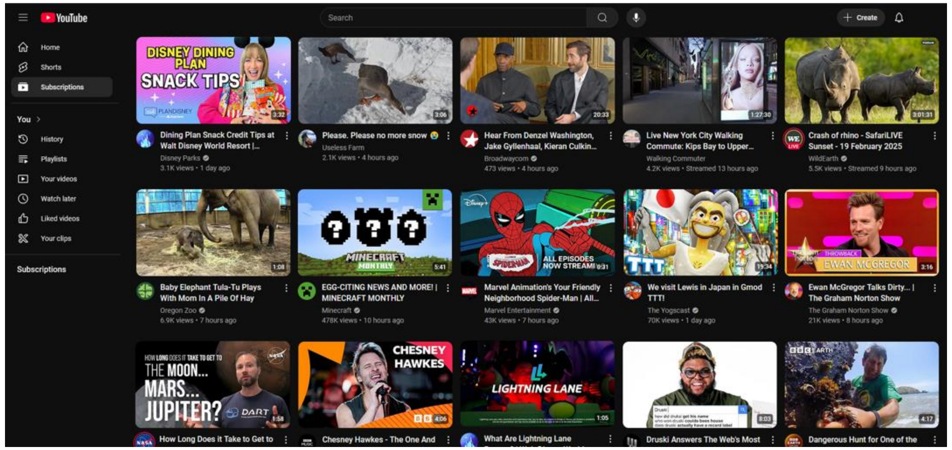
02 **Vision: generating from a learned space**

03 **Use case: image classification**

04 **Use case: recommendation systems**

05 **Challenge: design a responsible AI system**

Recommendation Systems

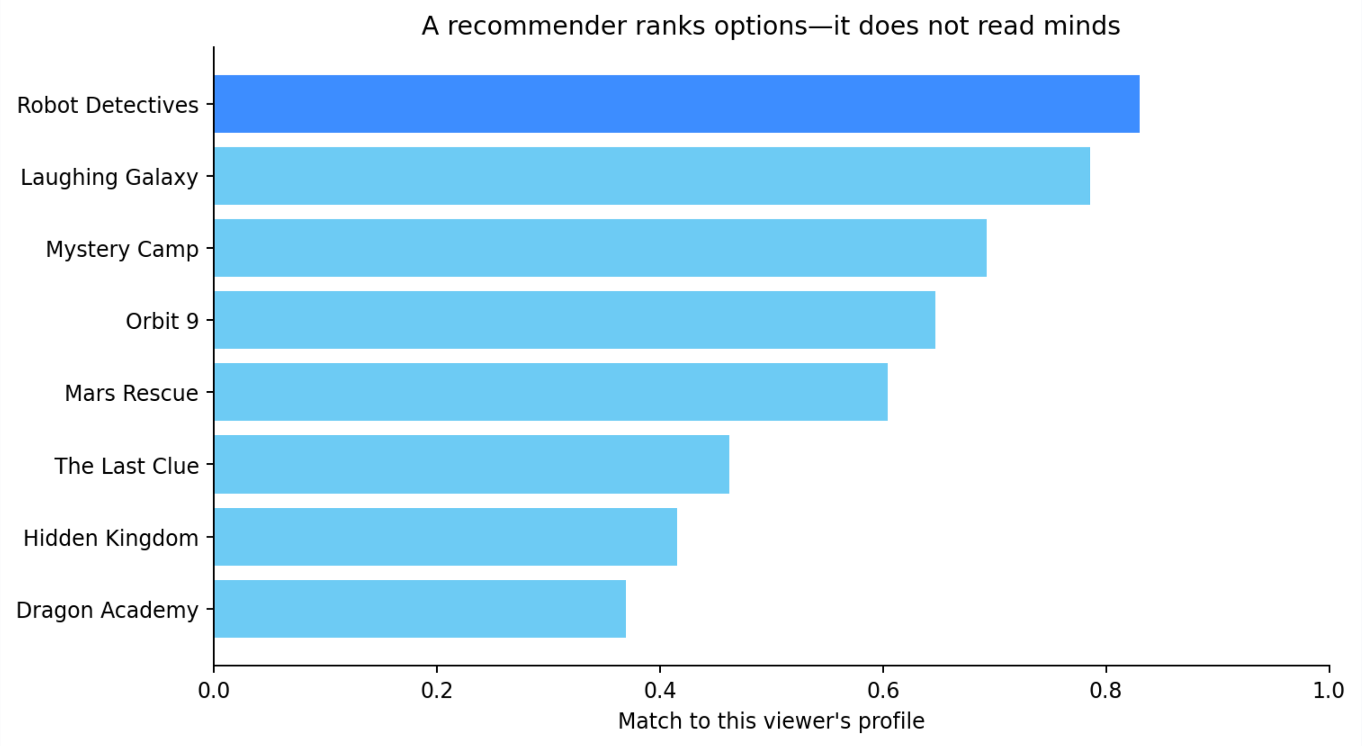


Recommendation is usually ranking

score each option for this user and context → sort → show a few

Signals may include content, past choices, similar users, freshness and business objectives.
The chosen objective shapes what people see.

Demo 4 · A transparent movie recommender



DEMO

Recommendations and feedback loops

Rank

The system selects what is visible

Click

People choose from what they saw

Learn

Clicks become new training signals

Narrow

Popular or familiar items gain more exposure

AI needs to grounding on evidence

Healthcare

Decision support, clinical validation, uncertainty and qualified review

Autonomous systems

Robust sensing, safe boundaries, simulation and emergency fallback

Creative tool

Disclosure, consent, content controls

Healthcare AI can support but human judgment still needed

Perceive

Highlight patterns in scans, signals or records for review.

Predict

Estimate risk with **uncertainty** and known limits.

Decide

A responsible professional combines evidence, context and patient values.

Thank you!