STIAEM

Slow Tech Innovation

AI, Ethics and New Teaching Methods

Connectionist Paradigms and Autonomous Extraction of Statistical Regularities

Co-funded by the European Union

2025-1-IT02-KA220-SCH-000365213

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them... Project reference code: 2025-1-IT02-KA220-SCH-000365213

Università di Foggia
Studi Umanistici, Lettere, Beni Culturali, Scienze della Formazione

STIAEM

Slow Tech Innovation
AI, Ethics and
New Teaching Methods

Co-funded by
the European Union

2025-1-IT02-KA220-SCH-000365213

eduOpen

CC BY-NC-SA 4.0 International

This license requires that reusers give credit to the creator. It allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, for noncommercial purposes only. If others modify or adapt the material, they must license the modified material under identical terms.

Università di Foggia
Studi Umanistici, Lettere, Beni Culturali,
Scienze della Formazione

OER 2.1.2: Connectionist Paradigms and Autonomous Extraction of Statistical Regularities
Reference Module: Module 2 - AI Fundamentals and LLM Mechanics
Related Video-Pill: Video 2.1 - Evolution of AI and Machine Learning
Authoring partner: University of Alicante – Spain
Topic / Thematic Focus: Connectionist paradigms, autonomous extraction and statistical regularities

Introduction and Learning Objectives

This resource completes, on the level of how things work, the path begun with OER 2.1.1 and accompanies video-pill 2.1 'Evolution of AI and Machine Learning'. If the previous resource reconstructed the history of the paradigms, this one focuses on the mechanism: how a neural network learns representations, what it means that it extracts them autonomously, and why all this has a direct impact on the way we teach students to read artificial intelligence critically.

The text connects to DigComp 3.0 (Cosgrove & Cachia, 2025) and to the 'pedagogy of AI' pillar of the UNESCO framework for teachers (Miao & Cukurova, 2024), precisely where the teacher is asked to understand where technology genuinely enhances learning, avoiding uncritical accelerationism.

By the end of the reading the teacher will be able to:

- describe what latent representations are and why they are organised hierarchically;
- explain the mechanism of attention (self-attention) and the advantage of parallelisation in Transformers;
- recognise in algorithmic models a statistical geometry rather than conceptual understanding, drawing didactic consequences from it.

For a historical overview of artificial intelligence, from its origins to today's systems, see Haenlein & Kaplan (2019).

1. Connectionism: units, weights, layers

The connectionist approach starts from a biological inspiration: knowledge is not represented by explicit rules, but distributed in a network of elementary units connected by connections of variable strength, the weights. A unit receives input signals, combines them according to its own weights and transmits an output; learning consists in progressively adjusting those weights, so that the network reduces the error between its answers and the expected ones.

A property follows from this that has important consequences on the didactic level: in the network there is no place where 'a cat has whiskers' is written. Knowledge is diffused throughout the whole system of weights, and no single unit can be interpreted as a rule. Already here one glimpses the theme, which we will encounter again in Module 3, of the opacity of systems: we can measure what the network does, we cannot always tell why it does it in that way.

2. Hierarchical latent representations

What does it mean, concretely, that a network 'learns a representation'? Let us imagine showing it the photograph of a face. The first layers, the closest to the input, respond to

simple elements: edges, contrasts, corners. The intermediate layers combine these elements into more complex components: an eye, a nose, a mouth. The deep layers put the components together and respond to the object as a whole: a face. No one designed this hierarchy from the outside: it emerges from training, layer by layer.
A simple exercise with students helps to make the difference visible: asking them to describe the steps by which they recognise a face in a photograph, and then comparing them with the hierarchy of layers. The conclusion reached is valuable: our perception also proceeds in stages, but the final result is a conscious experience endowed with meaning, whereas that of the network is a probabilistic classification devoid of experience.
It is worth highlighting a didactic consequence of this architecture. If the representation of an image is distributed over thousands of learned dimensions, there is no point in the network where the meaning of what is recognised is inscribed. Asking a model 'why did you classify it this way?' obtains a plausible explanation produced after the fact, not the reconstruction of the actual process: this is a point we will return to in Module 3, devoted to the opacity of systems.
These representations are called latent because they do not correspond to labels of human language, but to positions in a space of many dimensions - a vector space - learned from data. Every input is projected into that space, and the closeness between points expresses a similarity 'according to the model'. It is a useful and powerful geometry, but it is not a dictionary of concepts: it is the result of statistical regularities extracted from examples.
This organisation of knowledge on increasingly higher levels of abstraction is at the centre of the deep learning described by LeCun, Bengio & Hinton (2015).

3. The autonomous extraction of statistical regularities

Until the early 2000s, the most delicate step of machine learning was feature engineering: it was the human being who established which properties of the data counted. Deep learning reverses this relationship: it is the layers of the network that identify the relevant features by themselves, building increasingly abstract representations. It is this autonomous extraction of regularities that explains the leap in performance recorded from 2012 onwards, when deep convolutional networks markedly surpassed previous methods in image-recognition competitions.
The notion of embedding deserves an example. If in a space learned from data the words 'doctor' and 'hospital' turn out to be close, this does not mean that the model knows what a hospital is: it means that in the texts with which it was trained those words occur in similar contexts. Hence an operational criterion for the teacher: when a tool makes a mistake, the useful question is not only 'how do I correct it?', but 'on what data did it learn, and what do those data represent?'
The flip side of the coin is twofold. On the one hand, effectiveness is not accompanied by transparency: the learned representations have no meaning readable in natural language. On the other, the extraction of regularities depends entirely on the distribution of the training data: if that distribution is distorted, the learned regularity will be a distorted regularity, reproduced with statistical consistency. For the teacher, the consequence is that talking about data is not talking about a technical detail, but about a device that transfers into the system the prejudices present in the examples.
The definition of machine learning as the study of programmes that improve their effectiveness with experience is that of Jordan & Mitchell (2015).

4. Attention and the Transformer

It is worth clarifying that attention is not a form of understanding: the model does not know what a bank is, it knows that in certain contexts that representation is useful for predicting the next element of the sequence. The difference between 'using context' and 'understanding context' is exactly the difference that the school must make visible.

One last step remains, decisive for the language models. In previous architectures the sequence was processed one element after another, and as the distance grew so did the risk of losing the initial information. The Transformer, introduced by Vaswani and colleagues in 2017, abandons recurrence and relies entirely on attention.

«Self-attention, sometimes called intra-attention is an attention mechanism relating different positions of a single sequence in order to compute a representation of the sequence» (Vaswani et al., 2017, section 2).

In intuitive terms, to represent a word the model relates it to all the others in the sequence and weighs their mutual relevance: context is not a backdrop, it is what determines meaning. This is why the word 'bank', in the sentence 'the river bank', is represented differently from how it would be in 'the bank closed the account'. The technical advantage is parallelisation: not having to proceed in strict sequence, the computation exploits the hardware more efficiently, shortening training times and improving results. From this architecture descend the large language models, such as BERT and GPT.

5. Geometry, not concept: didactic implications

What should a teacher retain from this reconstruction? A simple and powerful idea: the model builds a geometric space learned from data, not a system of concepts. This means that its answers are plausible, not guaranteed; that its competence is a surface, statistical competence; and that its possible 'mistakes' are not lapses, but structural consequences of the mechanism. It is precisely from here that video-pill 2.2 begins, devoted to the probabilistic functioning of models and to the origins of hallucinations.

There is also an impact on the very idea of the tool's reliability: a model that builds representations rather than concepts cannot guarantee the correctness of what it produces, it can only make it statistically plausible. This explains why fact-checking remains indispensable and why it is didactically useful to ask students to trace the sources of generated content.

On the didactic level, understanding the mechanism opens up concrete activities. Students can be asked to describe how the meaning of a word changes as the context varies, to show what attention does; the hierarchy of layers can be reconstructed with visual examples, from pixel to object; one can discuss what 'similarity' means for a system that works on vector representations. These are exercises that do not require computer-science skills: they require the discipline of critical thinking.

Level of abstraction	What it learns	Visual example	Question to ask in class
Shallow layers	Simple elements	Edges, contrasts, corners	What does the machine see when it looks at an image?
Intermediate layers	Components	Eye, nose, mouth	How do elements combine into meaningful parts?
Deep layers	Complex objects	The face as a whole	Does the machine recognise or merely classify?

Attention (language)	Relations between positions of the sequence	'bank' with and without context	How does the meaning of a word change as the sentence varies?

6. Cognitive and Ethical Risks of Example-Based Prompting

Few-shot prompting can create an impression of reliability because the outputs become more predictable. However, predictability can generate false confidence. A response may look trustworthy because it resembles the examples while still containing factual inaccuracies or inappropriate assumptions.

The first major risk is cognitive offloading. If learners repeatedly delegate both the production and evaluation of answers to AI, the technology may reduce the amount of independent processing required from the learner. For this reason, educators should consider carefully which stages of an activity benefit from AI support and which stages require independent cognitive effort.

The second risk is imitation without understanding. This is particularly important when stable output structures are mistaken for evidence of conceptual learning. Learners may reproduce the visible form of a response without understanding the principles that produced it. A third risk concerns bias replication. Examples may contain cultural assumptions, stereotypes or narrow definitions of successful performance. Because few-shot prompting encourages the model to follow the demonstrated pattern, poorly selected examples may cause these limitations to be repeated across subsequent outputs.

A fourth risk is over-standardisation. Consistent formats can be useful for classification, feedback or structured learning materials. However, not every educational activity should lead to a uniform response. Tasks involving creativity, divergent thinking or personal expression may require more open-ended forms of AI interaction.

These risks reinforce the importance of DigCompEdu Area 6, which includes supporting learners in using digital technologies critically and responsibly. They also connect with the STIAEM concern that technological efficiency should not become more important than human judgment and cognitive development.

Visual Scaffolding

AI Mechanics for the Classroom:
From Statistical Geometry to Critical Thinking

This infographic explores how neural networks autonomously extract statistical regularities through connectionist weights and hierarchical layers. It highlights the distinction between a machine's "statistical geometry" and human conceptual understanding, providing teachers with a framework for AI pedagogy.

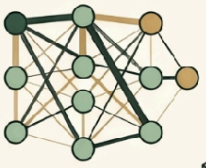
Inside the Network: How AI "Learns"

Knowledge as Distributed Weights

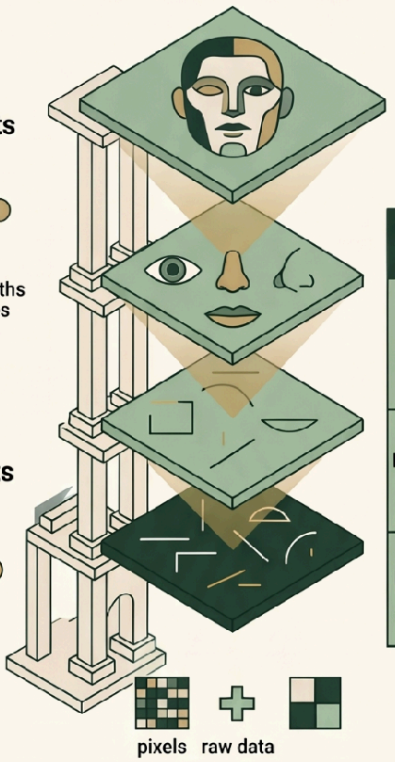


Variable connection strengths store knowledge, not rules

Knowledge as Distributed Weights



Variable connection strengths store knowledge, not rules



Attention vs. Understanding



self-attention
Transformers use context without without true understanding.

Abstraction Level	What it Learns	Question to Ask in Class
Shallow Layers	Edges, contrasts, and corners	What does the machine see when it looks at an image?
Intermediate Layers	Components (eye, nose, mouth)	How do elements combine into meaningful parts?
Deep Layers / Attention	Complex objects and contextual relations	Does the machine recognize or merely classify?

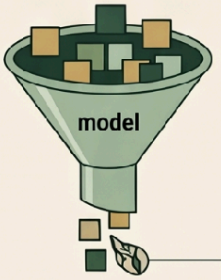
The Critical Classroom: Didactic Implications

Statistical Geometry, Not Concepts



AI responses are statistically plausible rather than guaranteed to be correct or meaningful.

Training Datasets



The Data Bias Trap

Models reproduce distorted regularities and prejudices found within their training datasets.



Models reproduce distorted reflections of societal biases.



Fact-Checking as a Necessity

Because AI competence is surface-level, students must trace sources and verify all generated content.



Gemini Notebook

Conclusions and Metacognitive Reflection Prompts

This resource has illustrated how connectionist networks work: units and weights, hierarchical latent representations, the autonomous extraction of statistical regularities, attention and parallelisation. The destination is not a technical competence, but a professional awareness: the systems we use build statistical geometries, not concepts; they depend on data and on their distortions; they produce plausible, not guaranteed, answers. Knowing this is the condition for not delegating to them what the school must cultivate: the critical thinking of students.

To transfer these core ideas into one's own practice, the following reflection prompts are proposed:

- Which activity in your subject would allow students to 'see' the difference between statistical representation and conceptual understanding?
- How would you explain self-attention with an example drawn from your subject, without using computer-science language?
- In what way does awareness of the opacity of learned representations change your way of using AI tools in teaching?

Technical Metadata Sheet

Metadata Field	Specification / Parameter Value
Identifier	OER 2.1.2
Title	Connectionist Paradigms and Autonomous Extraction of Statistical Regularities
Creator	University of Alicante - Spain
Description	The resource explains the mechanics of connectionist models: units, weights and layers, hierarchical latent representations and the autonomous extraction of statistical regularities. It describes self-attention and the parallelisation advantage of Transformers, with verified references (Vaswani et al., 2017). The didactic focus is on recognising that these models build a statistical geometry rather than conceptual understanding, and on drawing the consequences for teaching.
Keywords	connectionism; neural networks; latent representations; statistical regularities; self-attention; Transformer; feature learning; AI literacy
Language	EN (English)
Affiliation	University of Alicante - Spain
Contributor	STIAEM Transnational Scientific Board (peer review, Rhodes validation protocol)
Media Format	Text; Infographics
Genre	Training Manual; Educational Guide
URL	
Citation	Cosgrove, J., & Cachia, R. (2025). DigComp 3.0: European digital competence framework (Fifth edition). Publications Office of the European Union. https://data.europa.eu/doi/10.2760/0001149 LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444. https://www.nature.com/articles/nature14539 Miao, F., & Cukurova, M. (2024). AI competency framework for teachers. UNESCO. https://doi.org/10.54675/ZJTE2084 Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255-260. https://doi.org/10.1126/science.aaa8415 Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30. https://arxiv.org/abs/1706.03762
Educational Level	Secondary School

Associated Module / MOOC Video Lesson	Module 2 - AI Fundamentals and LLM Mechanics / Video 2.1 - Evolution of AI and Machine Learning
Learning Objectives	describe what latent representations are and why they are organised hierarchically; explain the mechanism of attention (self-attention) and the parallelisation advantage in Transformers; recognise in algorithmic models a statistical geometry rather than conceptual understanding, drawing the didactic consequences.
Rights	Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)
Date Available	14/09/2026
Date Modified	

References and Webography

Cosgrove, J., & Cachia, R. (2025). DigComp 3.0: European digital competence framework (Fifth edition). Publications Office of the European Union. https://data.europa.eu/doi/10.2760/0001149

Haenlein, M., & Kaplan, A. (2019). A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. Business Horizons. https://doi.org/10.1016/j.bushor.2018.08.004

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444. https://www.nature.com/articles/nature14539

Miao, F., & Cukurova, M. (2024). AI competency framework for teachers. UNESCO. https://doi.org/10.54675/ZJTE2084

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255-260. https://doi.org/10.1126/science.aaa8415

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30. https://arxiv.org/abs/1706.03762