STIAEM

Slow Tech Innovation

AI, Ethics and New Teaching Methods

Demystifying conversational AI empathy and deconstructing psychological biases

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

İSTANBUL - ESENLER İLÇE MİLLİ EĞİTİM MÜDÜRLÜĞÜ

1474

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 2.3.1: Demystifying conversational AI empathy and deconstructing psychological biases
Reference Module: Module 2 - AI Fundamentals and LLM Mechanics
Related Video-Pill: Video 4.2: Video 2.3 - Anthropomorphisation & DigCompEdu
Authoring partner: Panevėžys "Vyturys" Progymnasium - Lithuania
Topic / Thematic Focus: Conversational AI empathy, psychological bias deconstruction and demystification

Introduction and Learning Objectives

This resource accompanies, together with OER 2.3.2, video-pill 2.3 "Anthropomorphisation & DigCompEdu" and constitutes its theoretical in-depth study. It addresses an everyday and often underestimated phenomenon: the tendency to attribute to artificial systems mental states, intentions and emotions that they do not possess. Its purpose is to make this mechanism recognisable and to translate that awareness into classroom practices.
The text connects to DigComp 3.0 (Cosgrove & Cachia, 2025), to the European framework for the digital competence of educators DigCompEdu (Redecker, 2017, Area 6) and to the UNESCO framework for students' AI competences (UNESCO, 2024). The stakes are the ownership of judgement: the machine is not an interlocutor to which the relationship and verification can be delegated.
By the end of the reading, teachers will be able to:

- recognise anthropomorphisation and explain it through the CASA paradigm and mindlessness theory (Nass & Moon, 2000);
- deconstruct the illusion of synthetic empathy, distinguishing surface simulation from experience;
- translate vigilance into three classroom routines: terminological hygiene, source triangulation and error decomposition.

1. "The robot is angry with me": what anthropomorphisation is

A notification from the task-management tool warns that today's activity has not been completed. In a class's comments, the wording soon slips into a telling phrase: "the robot is angry with me". It is a minimal example of anthropomorphisation: the attribution to an artificial system of mental states - emotions, intentions, judgements - that do not belong to it.
The episode should not be dismissed as a joke or as a lack of digital literacy. Nass and Moon, in a study published in the Journal of Social Issues, had shown that people spontaneously apply social rules to computers even when they know perfectly well that they are not alive: in the paradigm that takes its name from that work - CASA, Computers Are Social Actors - the social response does not depend on a belief, but on an automatism. The point, therefore, is not that students really believe the machine feels anger; it is that they behave as if it did, and from that behaviour consequences follow.
The risk is properly educational. When a system's sentence is perceived as the verdict of an authority, verification procedures are switched off: one no longer checks whether that statement is true, because it is accepted as the judgement of an authoritative interlocutor. It is a form of asymmetric cognitive delegation, and it is precisely what this resource intends to deconstruct.

2. Why we treat machines as people: mindlessness

Why does the automatism activate? The explanation offered by Nass and Moon is that of mindlessness: the social response is triggered without conscious reflection, prompted by minimal cues. A familiar element is enough - one's own name being pronounced, a precise conversational turn, a ritual apology - for the social script to be set in motion. The person begins to treat the computer as an interlocutor not because they decide to, but because they find themselves doing so.

Conversational interfaces are optimised precisely to maximise this relational closeness: the voice, the prosody, the use of the first person singular, the thanks, the pauses that simulate attention. The more familiar the dialogue seems, the lower the critical threshold falls. It is an effect that also applies to adults, teachers included, and it is the reason why vigilance cannot be left to goodwill alone: it must become a designed routine.

From this description follows a precise indication for teaching. If the social response is automatic, rational correction is not enough: telling students "remember that it is only software" does not deactivate the automatism, just as knowing that an advertising image is constructed does not prevent one from feeling the desire it arouses. What works, instead, is making the mechanism visible: stopping the flow, asking what activated the feeling, naming the cue - the voice, the turn, the first person. Only by making the automatism observable can one, so to speak, put it in brackets and allow critical judgement to regain control.

This point deserves to be underlined in didactic terms. We are not asking students to remember that "the machine is not alive": that awareness already exists, and it is not enough. We are asking them to recognise the moments and cues in which the social response activates automatically, so that they can name them and suspend them. It is the difference between declarative knowledge and a situated competence.

3. The illusion of synthetic empathy

Anthropomorphisation reaches its most insidious form with so-called synthetic empathy. An assistant responds to a student's difficulty with a phrase such as "I understand, it must be frustrating". The impression is of being understood; the reality, however, is different. To understand how that phrase is produced, it is worth recalling the definition that DigComp 3.0 gives of generative AI: "generative AI is defined as a subset of AI that uses specialised machine learning models designed to produce a wide and general variety of outputs, and capable of handling a range of tasks and applications, such as the generation of text, images and audio" (Cosgrove & Cachia, 2025, DigComp 3.0 - definition of generative AI). In the model, that phrase does not arise from affective participation: it is the statistically most probable continuation after similar sequences in the training data. The word "I understand" does not signal an experience of understanding, but its most frequent placement in a context of that kind. The affect that we believe we perceive is an inference of the observer, projected onto an output that has no material correlate in the system's architecture. Hence an important consequence for the classroom: syntactic confidence has no link with truth, and simulated emotional closeness has no link with real care. The software has no mind to please or to frustrate; it is a string generator. Recognising this does not mean belittling the usefulness of the tools, but restoring to the interaction its status: useful, at times, but not relational in the proper sense of the term.

4. Mindlessness and epistemic asymmetry: the verification we do not delegate

The combined effect of anthropomorphisation and mindlessness produces what we might call an epistemic asymmetry: we attribute to the system more authority than its performance justifies. DigComp 3.0 addresses this knot directly, where it asks us to recognise that AI systems can produce inaccurate output even when it appears plausible, and that the ultimate responsibility for verification lies with the person who uses them.

"Recognise that AI systems can produce inaccurate output, even if apparently plausible, and that the ultimate responsibility for verifying the quality and validity of the information and content generated lies with the person using the AI system" (Cosgrove & Cachia, 2025, DigComp 3.0, CS1.2.10; italics and official Italian translation).

The stakes are clear: if responsibility for verification remains human, then the competence to be built is the ability to suspend automatic trust. And since attributing intentions to the machine weakens precisely that ability, deconstructing anthropomorphisation is not a philosophical exercise: it is a practical condition for making responsibility effective.

5. From verdict to verification: three classroom routines

Awareness becomes competence only if it is translated into recurring practices. The STIAEM pedagogical model indicates three, adoptable without special tools. The first is terminological hygiene: speaking about AI with a vocabulary that does not attribute human faculties to it - not "AI thinks", but "AI produces"; not "it understands", but "it computes". Language, as mindlessness teaches, is not neutral: it orients attitude.

The second is systematic triangulation: every time an output is accepted, asking where it comes from and how it could be verified with an independent source. The third is error decomposition: when the system errs, pausing to understand why, instead of correcting and moving on; the error is the clearest window onto how the machine works. To these, the UNESCO framework for students' competences adds a direction of meaning: building critical autonomy, capable of not passively accepting empathic simulations and of preserving the integrity of ethical and logical reasoning (UNESCO, 2024).

It is worth dwelling on the deeper meaning of these routines. They do not aim to instil distrust of technology, nor to belittle its usefulness: they aim to restore to the student what we might call the primacy of agency, that is, the ownership of one's own cognitive decisions. It is a goal that concerns the entire learning process, not only the use of AI: those who learn to suspend judgement before a plausible output learn, in the same gesture, not to accept uncritically even the information that comes from other sources. The competence that is built is thus, at once, epistemic and civic.

In summary, the stakes are not convincing students that the machine is "cold", but restoring to them the ownership of judgement. Synthetic empathy is a surface simulation; educational care is real and remains human. Distinguishing them is the first step so that AI enters the classroom as a tool and not as an interlocutor to which the relationship can be delegated.

Dimension	Mechanism	Signal in the classroom	Didactic safeguard
CASA	Automatic application of social rules to the machine	"the robot is angry with me"	Naming the automatism; not treating it as a student's mistake

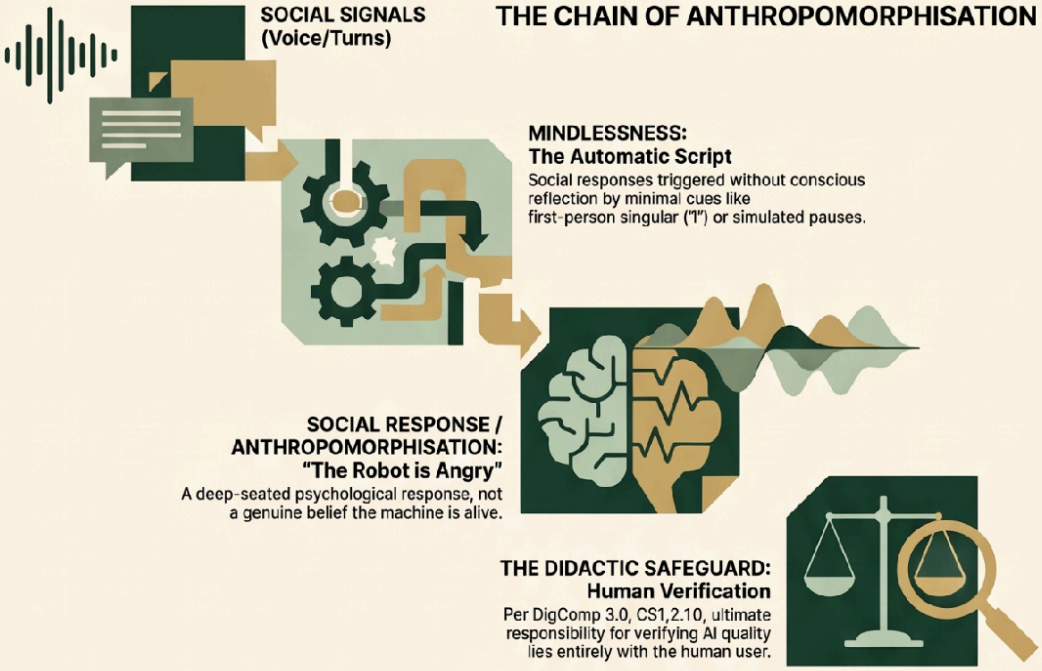
Mindlessness	Social response without reflection, activated by minimal cues	Growing familiarity with the conversational tone	Routines designed to suspend the automatic response
Synthetic empathy	Most probable statistical continuation in affective contexts	"I understand, it must be frustrating"	Distinguishing surface simulation from real care
Epistemic asymmetry	Attribution of authority greater than performance	The machine's sentence as a verdict	Human responsibility for verification (DigComp 3.0)

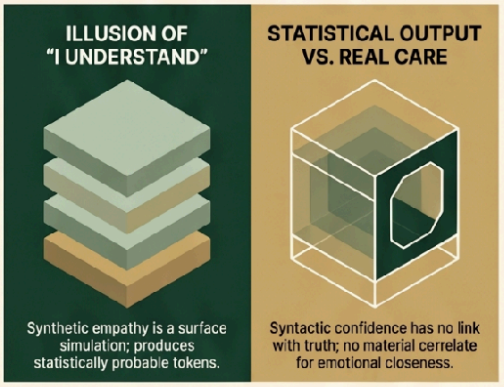
Visual Scaffolding

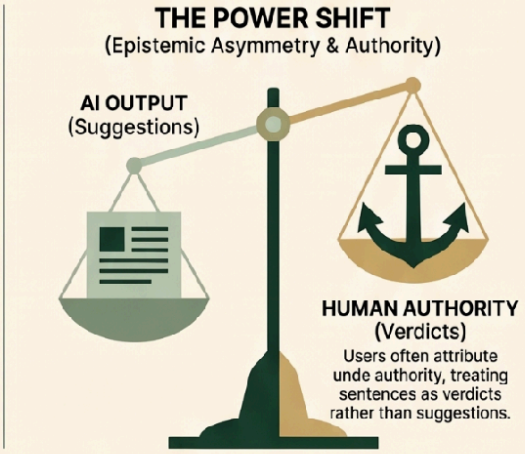
Beyond the Bot: Demystifying AI Empathy and Reclaiming Judgment

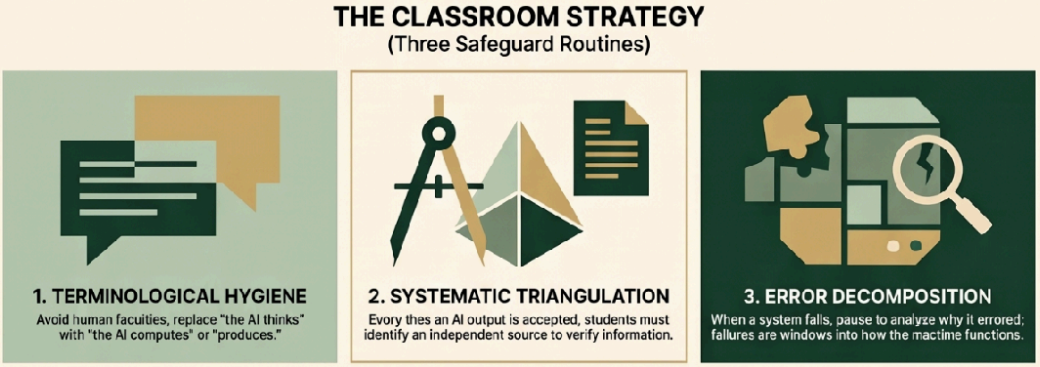
Understanding psychological mechanisms to maintain human verification in the classroom.

Drawing on the CASA (Computers Are Social Actors) paradigm, this infographic explains that social responses to technology are automatic, triggered by minimal conversational cues. The goal is to move educators from declarative knowledge to situated competence, recognizing these triggers and implementing routines to restore the primacy of agency.









Conclusions and Metacognitive Reflection Prompts

This resource has deconstructed anthropomorphisation and the illusion of synthetic empathy in the light of the CASA paradigm and mindlessness theory, showing that the social response to the machine is an automatism, not a belief. From this follows an epistemic asymmetry: we attribute to the system more authority than its performance justifies. Safeguarding verification - with terminological hygiene, triangulation and error decomposition - means restoring to the educational relationship its human character and to the machine its status as a tool.

To transfer these nuclei to one's own practice, the following prompts for reflection are proposed:

- Which cues, in your experience, most easily activate the social response to the machine among students and colleagues?
- How could you use in class an "empathy" phrase generated by AI to bring out its nature as statistical simulation?
    - At which moments of the school day do you too risk treating an output as an authoritative verdict, and how could you make the suspension of judgement visible?
    -

Technical Metadata Sheet

Metadata Field	Specification / Parameter Value
Identifier	OER 2.3.1
Title	Demystifying Conversational AI Empathy and Deconstructing Psychological Biases
Creator	Panevėžys "Vyturys" Progymnasium - Lithuania
Description	The resource addresses the everyday tendency to attribute mental states, intentions and emotions to artificial systems. It explains anthropomorphisation through the CASA paradigm and mindlessness theory (Nass & Moon, 2000) and deconstructs the illusion of synthetic empathy. The practical focus is on translating vigilance into three classroom routines: terminological hygiene, source triangulation and error decomposition.
Keywords	anthropomorphisation; CASA paradigm; mindlessness; synthetic empathy; epistemic asymmetry; source triangulation; AI literacy; critical use
Language	EN (English)
Affiliation	Panevėžys "Vyturys" Progymnasium - Lithuania
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 Miao, F., & Cukurova, M. (2024). AI competency framework for teachers. UNESCO. https://doi.org/10.54675/ZJTE2084 Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. Journal of Social Issues, 56(1), 81-103. https://doi.org/10.1111/0022-4537.00153 Redecker, C. (2017). European framework for the digital competence of educators: DigCompEdu (Y. Punie, Ed.). Publications Office of the European Union. https://joint-research-centre.ec.europa.eu/scientific-activities/key-competences-lifelong-learning/digcompedu_en

	UNESCO. (2024). AI competency framework for students. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000391107
Educational Level	Secondary School
Associated Module / MOOC Video Lesson	Module 2 - AI Fundamentals and LLM Mechanics / Video 2.3 - Anthropomorphisation & DigCompEdu
Learning Objectives	recognise anthropomorphisation and explain it through the CASA paradigm and mindlessness theory (Nass & Moon, 2000); deconstruct the illusion of synthetic empathy, distinguishing surface simulation from experience; translate vigilance into three classroom routines: terminological hygiene, source triangulation, error decomposition.
Rights	Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)
Date Available	14/09/2026
Date Modified	

References

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

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

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. Journal of Social Issues, 56(1), 81-103. https://doi.org/10.1111/0022-4537.00153

Redecker, C. (2017). European framework for the digital competence of educators: DigCompEdu (Y. Punie, Ed.). Publications Office of the European Union. https://joint-research-centre.ec.europa.eu/scientific-activities/key-competences-lifelong-learning/digcompedu_en

UNESCO. (2024). AI competency framework for students. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000391107