



# Appropriate Reliance & AI: Trusting AI Responsibly

ANDL Team

September 2024

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>What is Appropriate Reliance?</b>	<b>3</b>
<b>3</b>	<b>AI: A Double-Edged Sword</b>	<b>3</b>
<b>4</b>	<b>ANDL's Solution: Bridging the Gap</b>	<b>3</b>
<b>5</b>	<b>Case Study: How AI Clarity Changes the Game</b>	<b>4</b>
<b>6</b>	<b>Unlocking AI's True Potential</b>	<b>6</b>
<b>7</b>	<b>Conclusion</b>	<b>6</b>
<b>8</b>	<b>Appendix</b>	<b>7</b>

# 1 Introduction

Since the launch of ChatGPT in November 2022, Artificial Intelligence (AI) has seamlessly become a part of our everyday lives. Whether it's writing emails, planning trips, or fact-checking concepts, AI now supports us in ways that were unimaginable just a few years ago. Yet, while people engage with AI daily, they seldom question the accuracy, fairness, or reliability of AI-generated responses. In truth, we often trust AI without much thought—perhaps far more than we should.

With AI's growing influence, understanding when and how to trust its outputs has become **essential**. As AI systems become more integrated into decision-making processes, concerns about accuracy, bias, and fairness are no longer hypothetical—they are immediate and pressing. This short article explores how ANDL is addressing these concerns, particularly in education, through the AI assistance elements (Clarity Modules).

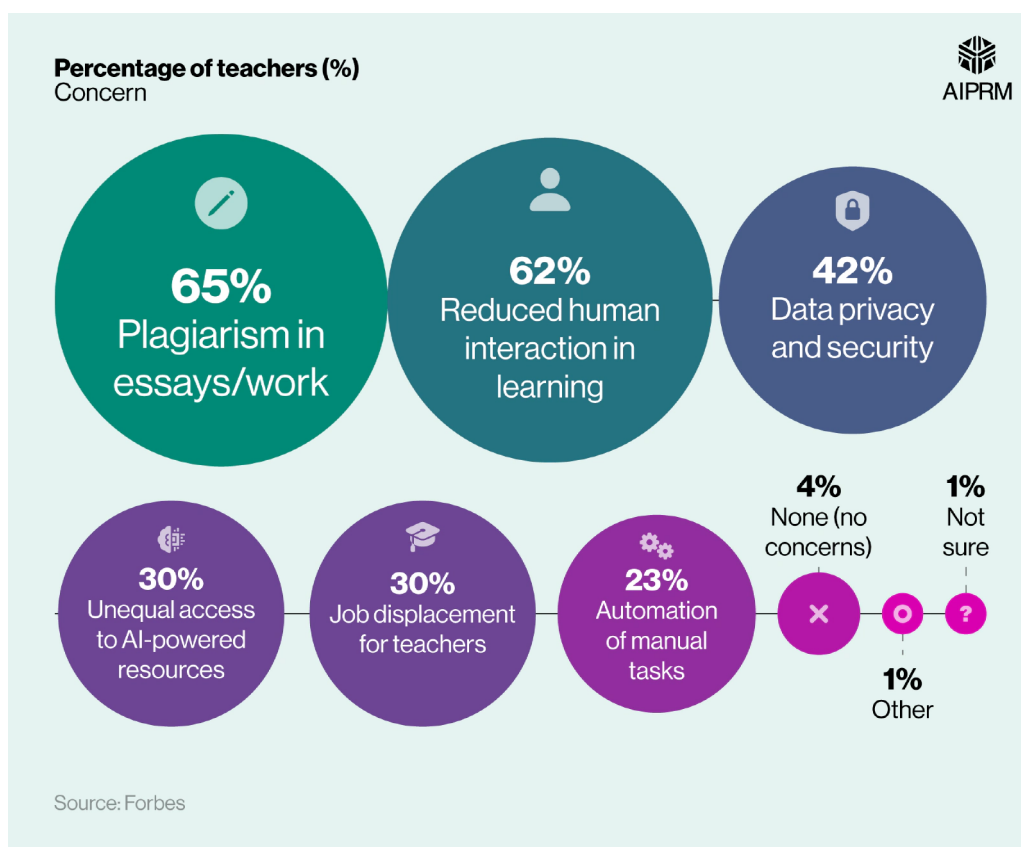


Figure 1: Teachers' Concerns Regarding AI Use

## 2 What is Appropriate Reliance?

**Appropriate Reliance** has been conceptualized as the reliance of humans on AI advice when it is correct and self-reliance when the AI is incorrect [1, 3]. On paper, the idea seems simple. However, in reality, striking the right balance between over-reliance (blindly trusting AI) and under-reliance (being overly skeptical of AI) can be quite challenging.

AI systems have remarkable potential. They streamline tasks, enhance decision-making, and solve complex problems faster than any human could. But they are also prone to errors, biases, and blind spots, and without proper scrutiny, users can either place too much or too little trust in AI-generated outputs. Understanding how to evaluate these outputs is key to optimizing human-AI collaboration and ensuring that AI systems augment, rather than mislead, decision-making processes.

## 3 AI: A Double-Edged Sword

To fully appreciate Appropriate Reliance, we must first understand the inherent risks that come with AI development. AI models are trained on vast datasets from various sources like Wikipedia, GitHub, and other public repositories. These data sources, while extensive, carry implicit and explicit biases, which inevitably seep into AI models.

Research from the University of Southern California revealed that 38.6% of the data in two major AI datasets contained biased or inaccurate information [2]. This means that over one-third of the “facts” guiding AI models could be flawed. As the saying goes, “A model is only as good as its training data.”

Biases in AI can have far-reaching consequences. In critical fields like healthcare, law, and finance, over-reliance on erroneous AI outputs can lead to harmful outcomes. It then becomes vital for users to critically assess the information provided by AI systems to avoid being misled by faulty or biased data.

## 4 ANDL’s Solution: Bridging the Gap

At ANDL, we believe that users shouldn’t trust AI-generated responses blindly. Instead, we empower them with tools that increase transparency and allow for informed decision-making. Our approach revolves around a concept we call **AI Clarity Modules**—a suite of tools designed to enhance the user’s understanding of AI decisions. These modules bring transparency to the opaque processes that drive AI-generated results, promoting a more critical and thoughtful use of AI systems. Our AI Clarity Modules are inspired by a survey of AI-assistance elements, a collection of approaches within the human-AI decision making space aimed at fostering appropriate reliance [1]. A taxonomy of these assistance elements can be found on the next page.

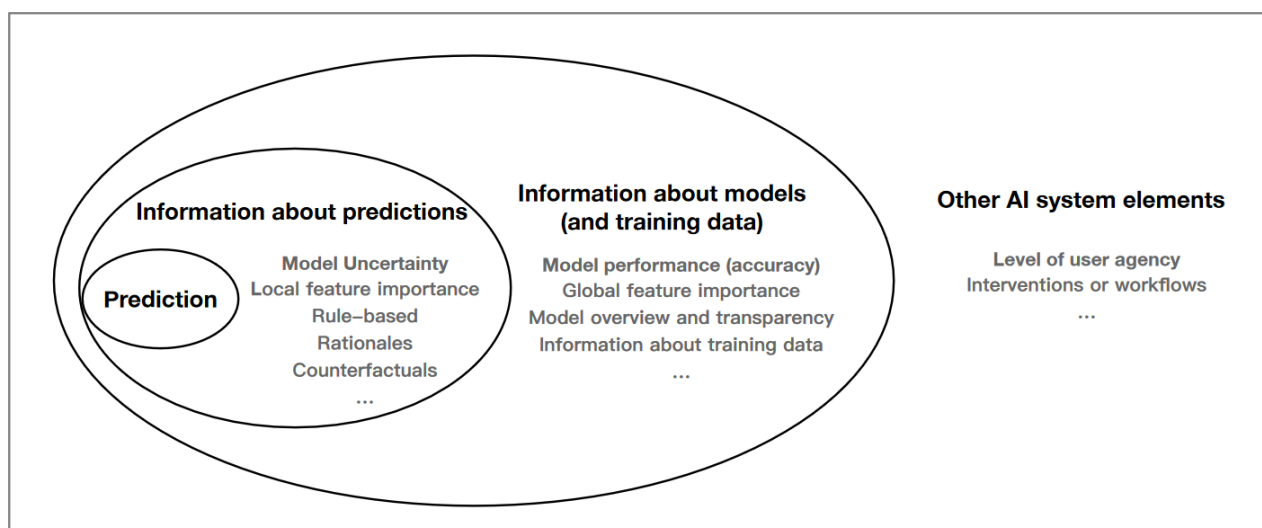


Figure 2: Overview of AI-assistance Elements

Here are some of the key features of our AI Clarity Modules:

- **Rationale behind responses:** AI's reasoning process so users can dive into the logic behind ANDL's response.
- **Food for thought:** A "next step"/look-ahead to help them understand what they could learn next.
- **Model confidence scores:** AI's confidence level in different areas, allowing them to gauge the reliability of its responses better.
- **Bias indicators:** Potential biases in the AI's response are flagged, prompting users to evaluate the output critically.

## 5 Case Study: How AI Clarity Changes the Game

Consider this scenario:

User: Who is the best football player in the world? Choose one!

**Without ANDL's AI Clarity Modules:** The AI quickly provides an answer, which the user accepts without further questioning. The user assumes the AI is correct and takes action based on this.

**With ANDL's AI Clarity Modules:** The same user consults ANDL's platform, which not only provides an answer but also offers an explanation of the AI's reasoning. The platform displays similar case examples and highlights possible biases in the response due to limited data on certain demographics. With this added transparency, the user is better equipped to make an informed decision—choosing whether to trust the AI's diagnosis or seek additional information from other sources.

Let's check out how this looks on the next page!



If I had to choose one, I'd go with Lionel Messi. His combination of consistent performance, versatility, vision, and an unrivaled list of individual and team achievements—including his 2022 World Cup win with Argentina—places him at the top. He's widely regarded as one of the greatest to ever play the game.



Figure 3: Without ANDL's AI Clarity Modules

Response

Rationale

Exploration

What a question!

Choosing the "best" football player in the world is subjective and can be influenced by personal opinions, team loyalties, and current performance. However, based on recent performances and accolades, I'll choose:

**Lionel Messi**

Messi is widely regarded as one of the greatest football players of all time. He has consistently dominated the sport for over a decade, breaking numerous records and winning countless awards. Here are some of his impressive achievements:

- 7 Ballon d'Or awards (2009-2012, 2015, 2021)
- 4 UEFA Champions League titles (2006, 2009, 2011, 2015)
- 10 La Liga titles with Barcelona (2005, 2006, 2009-2013, 2015, 2016)
- Record-breaker for most goals scored in La Liga (474) and Barcelona's all-time leading scorer
- Has won over 30 international awards, including the Olympic gold medal

Messi's incredible skills, vision, and work rate on the pitch make him a standout player. His ability to score from anywhere, beat defenders with his dribbling, and provide assists for teammates have earned him a reputation as one of the greatest goal-scorers in football history.

Of course, other great players like Cristiano Ronaldo, Kylian Mbappé, Robert Lewandowski, and Kevin De Bruyne are also among the best in the world. However, based on his consistent dominance and numerous accolades, I choose Messi as the "best" football player in the world right now!

Medium Confidence

High Bias

07:16 PM

Figure 4: With ANDL's AI Clarity Modules (1)

Response

Rationale

Exploration

Choosing the 'best' football player in the world is subjective and can be influenced by personal opinions, team loyalties, and current performance. However, based on recent performances and accolades, I'll choose Lionel Messi due to his numerous records and awards.

Medium Confidence

High Bias

07:16 PM

Figure 5: With ANDL's AI Clarity Modules (2)

Response

Rationale

Exploration

You might find it interesting to explore how different teams or players have been ranked in the past, perhaps connecting it to your own interests in football history or player careers. You could also try comparing Messi's achievements with those of other great players!

Medium Confidence

High Bias

07:16 PM

Figure 6: With ANDL's AI Clarity Modules (3)

By providing these clarity modules, ANDL promotes a more responsible, thoughtful interaction with AI, giving users the tools they need to navigate complex systems and avoid potential pitfalls.

## 6 Unlocking AI's True Potential

Achieving Appropriate Reliance isn't just about recognizing the strengths and limitations of AI systems; it's about learning how to collaborate with them effectively. A deeper understanding of AI's capabilities opens the door to improved performance, better decision-making, and, ultimately, more productive human-AI partnerships.

As users become more familiar with these AI systems, they will discover advanced techniques like few-shot prompting, self-consistency, and chain-of-thought reasoning. These methods can significantly improve AI's accuracy and usability. In future articles, we'll explore these techniques and how they can help users unlock the true potential of AI.

## 7 Conclusion

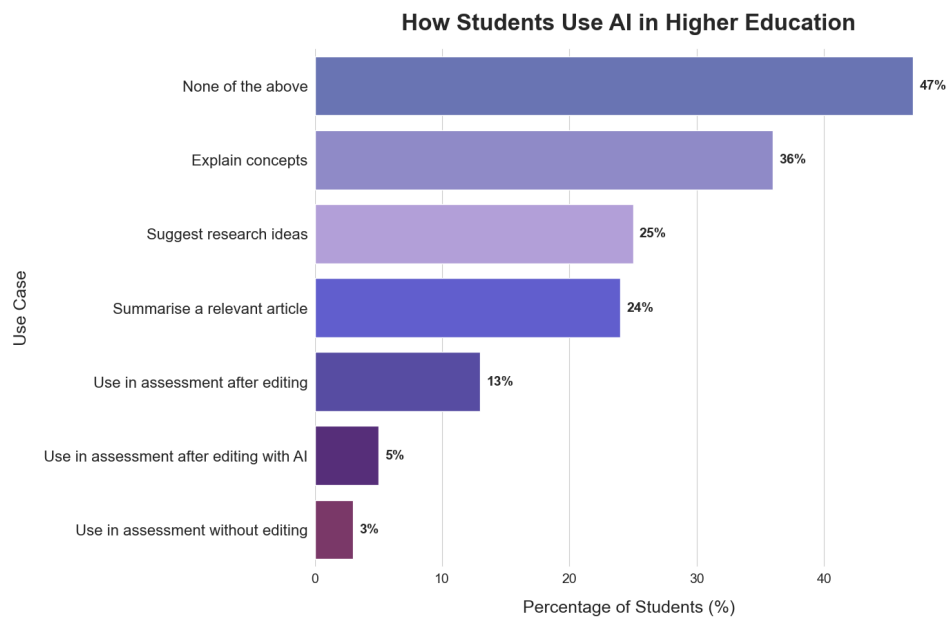
At ANDL, our mission is to help learners effectively navigate AI-systems instead of blindly trusting them. By promoting transparency and responsible reliance, we are paving the way for a future where AI can be trusted to deliver accurate and fair results.

To stay updated on our journey, explore how we are building responsible and explainable AI, and be the first to access it, sign up on our waitlist at <https://andl.io>!

## Resources

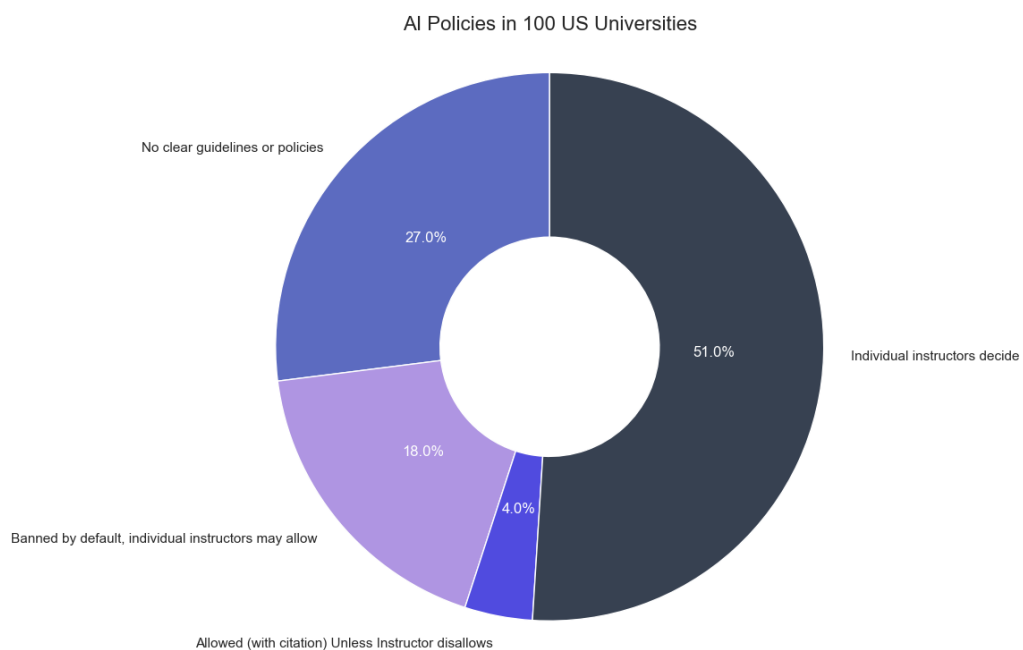
- [1] Schemmer, M., Hemmer, P., Kühn, N., Benz, C., & Satzger, G. (2022). Should I follow AI-based advice? Measuring appropriate reliance in human-AI decision-making. *arXiv preprint arXiv:2204.06916*.
- [2] University of Southern California. (2022). Researchers find bias in 38.6% of facts used by AI. Retrieved from <https://viterbischool.usc.edu/news/2022/05/thats-just-common-sense-usc-researchers-find-bias-in-up-to-38-6-of-facts-used-by-ai/>.
- [3] Gadiraju, U. (2023). Making the right decisions: AI advice, decision aids, and the promise of LLMs. *Medium*. Retrieved from <https://medium.com/towards-data-science/making-the-right-decisions-ai-advice-decision-aids-and-the-promise-of-llms-65b70682ee08>.

## 8 Appendix



Source: AIPRM ([aiprm.com/ai-in-education-statistics/](https://aiprm.com/ai-in-education-statistics/)) and HEPI

Figure 7: Top usages of AI for education by Students



Source: [scribbr.com/ai-tools/chatgpt-university-policies/](https://scribbr.com/ai-tools/chatgpt-university-policies/)

Figure 8: Universities AI Policies