AI Conceptualization and Policy Permissibility

Zea Miller, **Kashish Sachdeva, and Jake Walker**
The University of Florida

Abstract

When universities create AI policies, they often conceptualize AI as *something*, such as a tool or a resource. This study questions whether such policies are affected by how they envision AI. In other words, is the permissible a function of conceptualization? To answer, R1 university policies were rated independently by three raters on two axes: conceptualization and permissibility. When visualized, the ratings clearly show that while AI qua TOOL does not inherently attach either to the restrive or permissive, AI qua RESOURCE does not attach to the restrictive. Ultimately, this study shows that universities are unlikely at this time to conceptualize AI as a resource and simultaneously ban it.

Keywords: Artificial Intelligence; AI; Concept; Framing; Higher Education; Large Language Models; LLM; Policy; University

1 Introduction

Since the late 2022 arrival, or at least the broad availability, of large language models (LLMs), universities and their faculties have had to determine whether to adopt an artificial intelligence (AI) policy. In so doing, they might grapple with what an AI policy should be or what form it should take, what boundaries it introduces, and to whom it attaches.

Along the way, universities are increasingly responding to AI through policy creation and revision (Robert, McCormack 2025) but not many "have moved beyond reactive, classroom-based rules" (Legatt 2025). Some find it difficult "to keep up with the changing norms" (Hill 2025). Others, though, stand ready "to make A.I. tools part of students' everyday experiences" as part of new collaborations between universities and AI companies (Singer 2025).

How to respond to AI has been a prolific topic online, to say nothing of the conversations happening on campus, in webinars, during conferences, and in publications. What is emerging is an AI policy landscape *shaped* by the *language* of these conversations.

There is an immediate issue introduced by and through AI: when deciding what to do about AI on campus, to whatever extent, universities do not aim, for example, to ban self-driving cars or facial recognition; rather, they mean to police LLMs. Even so, LLMs are sometimes referred to in policy statements by popular platform names instead of their generic form. Called AI, tools, and at times resources, metaphoric positioning pervades when discussing LLMs, which not only influences our language but also our policy.

Since "ontological metaphors make concepts real" (Murray-Rust et al. 2022, 1), metaphors matter materially. AI, either as a multiplicity of things or in the simplicity of a TOOL as one thing, cannot be realized (that is, made real) in the conversation when "understanding the function of a tool is an essential step in learning to use a tool" (Menz et al. 2010, 1438) and banning use prevents such learning. AI is too many things to be simply understood as *something*, to say nothing of misunderstanding LLMs as one thing or the very processes behind how they function. Resources, however, need not be used to be understood (i.e., one need not go to a tutoring center to understand its function or benefit). One goes to a RESOURCE when it is needed and not when not. Resources abound on college campuses, so calling online resources like AI a resource in the same way one calls digital information one is not a stretch. By calling AI

itself AI or a TOOL or a RESOURCE, we are metaphorically cementing the semantics for an entire constellation of concepts and activating the "relationship with concepts situated on the same plane" where "concepts link up with each other, support one another, coordinate their contours, articulate their respective problems, and belong to the same philosophy, even if they have different histories" (Deleuze, Guattari 1994, 18). As follows, metaphors instantly activate conceptual connections that matter meaningfully, among which AI itself "relies on metaphors to construct the links between human thought, mathematical properties and the exigencies of technological possibility. The very term 'artificial intelligence' is a deeply burdened metaphor" (Murray-Rust et al. 2022, 3). After AI itself, the secondary use of metaphor (that is, calling the metaphor AI another metaphor [e.g., either tools or resources]) carries semantic, policy, and power considerations, inasmuch as "one of the most important avenues of semantic analysis has been the examination of figurative language, especially metaphor, in policy thinking" where "policy is the deployment of language for strategic purposes" (Kochis 2005, 28–29) and for which "discourse adopted within...policy communities becomes a crucial influence on the generation of policy" (Jacobs, Manzi 1996, 550–551).

Thus, conceptualization matters at the policy level semantically. As follows, policy formulation and framing should entail careful semantic calibration. For AI qua LLMs, what is found across higher education is a variety of AI conceptualizations and range of permissions. Yet, they share languaging features like metaphoric concepts and permissions and lead to shared conditions for which commonalities might be mapped. To that end, this study examines whether there is a correlation between how universities conceptualize AI (i.e., TOOL vs. RESOURCE) and how permissive they are toward AI (i.e., restrictive, balanced, or permissive). This study aims to answer the following question: what is being restricted vs. permitted? We wondered whether TOOL might attach to the restrictive in the same way that RESOURCE might attach to the permissive. Since universities and their faculties will be crafting if not revising their policies over time, the significance of this study rests in its informative power to

capture AI policy today and possibly influence AI policy framing tomorrow through semantics.

2 Methods

Since the optimal way to show whether a correlation between conceptualization and permissibility would be by and through the rating, tabulation, visualization, and analysis of data, the procedure followed the same structure. The study focuses solely on R1 universities based on the 2021 classification (n=146), not the 2025 revised list (n=187), which was released (spring 2025 term) after the analysis began (fall 2024 term). The focus on R1 universities stems from the idea that they would, as leading institutions with the most students and capacity for and access to AI technologies, formulate AI policies first.

2.1 Structuring Data

At the preliminary stage, plots were created using randomly-generated data to determine what models would be possible and how various rating scales would influence them. What fit best showed that the process would require two metrics with specific rating scales for each.

A list of all R1 universities in the United States (n=146) was set in a spreadsheet alphabetically. Initial columns tracked public or private status, city, state, and website where the AI policy could be found. Columns for rating both conceptualization (1, 5, 10) and permissibility (1–10) per university were allocated to each rater (n=3). Additional columns were added later for averaging ratings, rounding averaged ratings, and cluster (bin) tracing.

2.2 Rating Universities

To plot ratings of how universities conceptualize AI against how universities restrict or permit by and through their AI policies, a system of biaxial ratings were required.

For the conceptual axis, the authors established 3 ratings (mirroring the 1–10 scope used next): 1, 5, and 10 for TOOL, a combination of these terms, and RESOURCE or similarly not TOOL (e.g., LLMs, GPTs, etc.), respectively.

For the permissive axis, the authors established a scale of ratings ranging from 1-10: Radically Restrictive (1), Highly Restrictive (2), Strict (3), Structured (4), Balanced (5), Flexible (6), Encouraging (7), Empowering (8), Highly Permissive (9), and Radically Permissive (10).

Through an independent, unsupervised, and unalterable rating process, each rater scored how each university conceptualized AI and how permissive each university policy was toward AI use, generally, during the fall of 2024 to the spring of 2025. A second search was conducted during the spring of 2025 to determine whether any new policies emerged since the process began, and if so, the second scores for new policies were retained.

2.3 Tabulating, Visualizing, and Analyzing Data

All ratings were added to the group spreadsheet separately, and they remained unchanged after viewing (i.e., no rater changed their scores after viewing other scores). Additional columns (2) per rating group (conceptualization and permissibility) were allocated for averaging the first three in each (1) and rounding the result (2). Clusters were calculated for below (<5), at (=5), and above (>5) the rating midpoint value for permissibility. Likewise, conceptual clusters were determined for below, at, and above the rating midpoint values. Data were imported into Tableau and analyzed therein for speed. In time, plots were created in R using ggplot2 for visualization here. Finally, data were analyzed in R using statistical models.

3 Results

3.1 Data Visualizations

Processed in R using the package ggplot2, the data led to the following visualizations and ground further analysis.

3.1.1 Ratings by Institution Count

By plotting the count of institutions by their ratings, the intercepts of the conceptualization slope and the permissibility slope reveal that as conceptualization moves from tool to resource and as permissibility moves from restricted to permissive, the response to AI tends to be restricted before, balanced between, and permissive after.

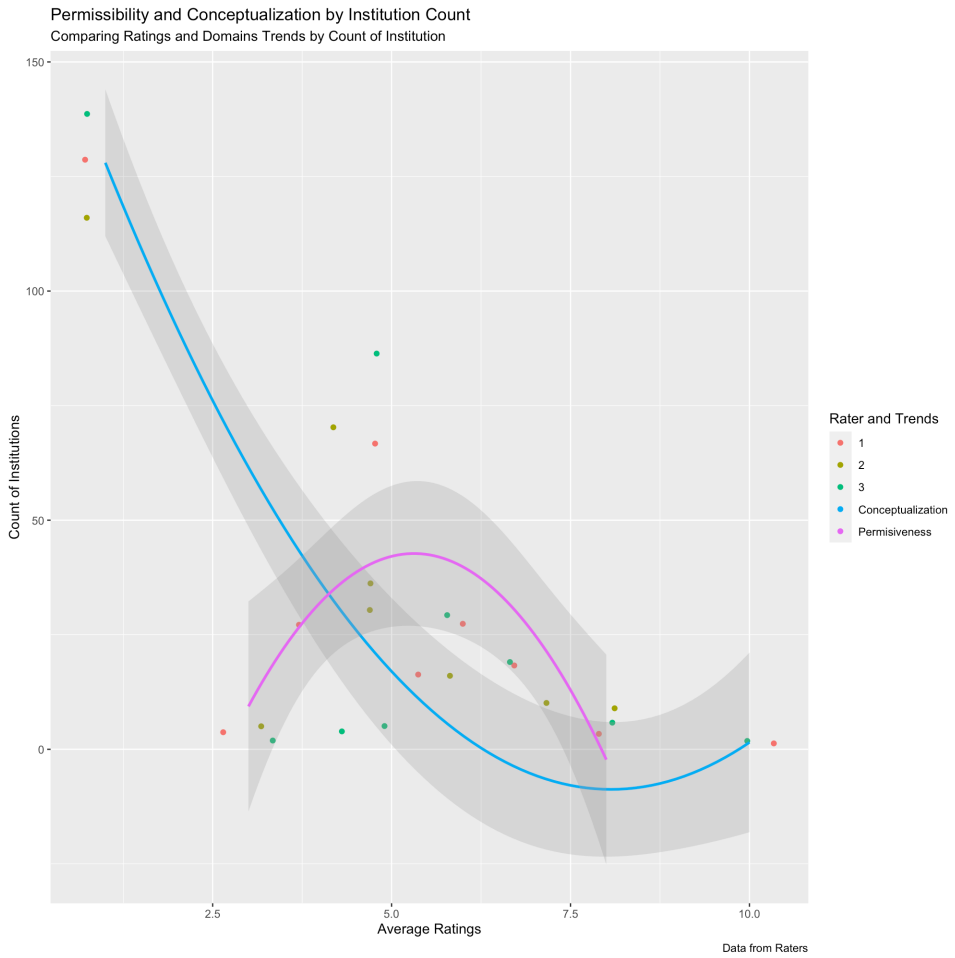

Figure 1: Intercepting domain slopes of conceptualization and permissibility by count of institution.

The count of institutions across both conceptualization and permissibility domains have slope intercepts that align with the conceptualization-permissibility gradient. Before the intercept, we are more likely to find TOOLS and restrictive. Where overlapping, we are more likely to find combination and balanced. After the intercept, we are more likely to find RESOURCE and permissive. Since Figure 1 uses counts of institutions with 9 intersecting potentials, it anticipates the 3x3 matrix of Figure 3.

3.1.2 Rating by Rating

Attempting to plot conceptualization against permissibility as raw ratings does not yield a trend because they have discreet and continuous data. Nevertheless, we see stratification, as expected.

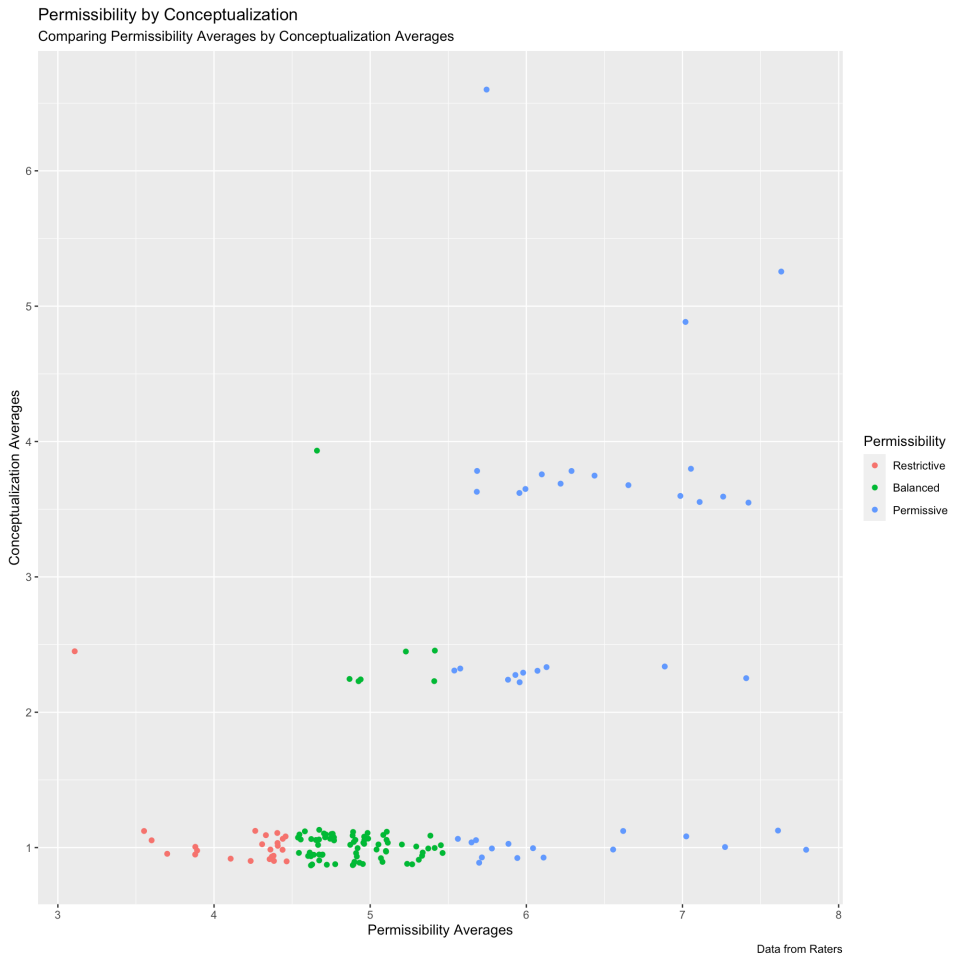

Figure 2: Stratification of the data.

We also see the loss of the restrictive over rising strata, which shows that tools attach to the entire policy spectrum while resource does not.

3.1.3 Clustering Ratings for Density

By visualizing the segmentation through institutional density, significant clusters (shades of blue) and voids (gray) become clear.

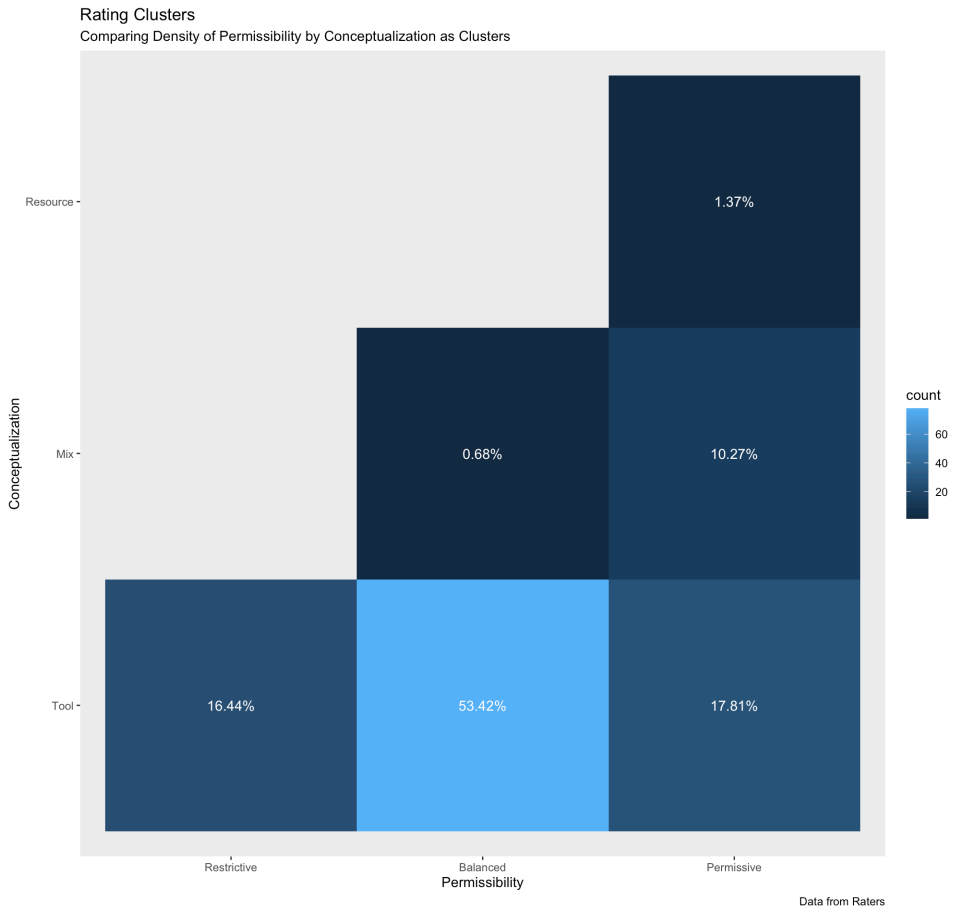

Figure 3: Clusters and voids in the data.

The intersections of conceptualization strata with the policy spectrum show presence by color and absence by lack. Without representatives in Restrictive-Resource or Balanced-Resource, resources are not restricted.

3.2 Statistical Models

Processed in R, the data return following outputs that ground further analysis.

3.2.1 Rater Agreement

To test the inter-rater reliability across the categories, Krippendorff's alpha was calculated. The result shows low agreement for both Permissibility (α = 0.379) and Conceptualization (α = 0.277).

Measure	α	Confidence Interval
Permissibility	0.379	0.256, 0.491
Conceptualization	0.277	0.162, 0.375

Table 1: Krippendorff's alpha for the inter-rater reliability of both permissibility and conceptualization scores.

The raters for permissibility seem slight to fairly aligned, with 95% confidence that the level of agreement would be between slight (0.256) to fair (0.491). The raters for conceptualization are slightly aligned, with 95% confidence that the level of agreement would be between low (0.162) to slight (0.375). As non-zero intervals, the observed agreement is significant but admittedly low, which is sufficient for one-shot, subjective, and exploratory research. However, low alignment on these fronts in Table 1 is easily attributed to, on the one hand, the infrequent instance of resource (n=3) for conceptualization and, on the other hand, the frequent agreement on permissibility (in that, the rarity of a rating or too frequent an agreement depresses the α coefficient). Since Krippendorff's alpha is a chance-corrected measure, and since it is suppressed by both the rarity of resource and commonality of rater agreement, comparing it to direct observations provides a necessary contrast: at least 2 of the raters agreed on conceptualization 96 out of 146 institutions, corresponding to an agreement frequency of 65.8%, which is moderate.

3.2.2 The Rating System

Since resource was a rare category in the data, and since the raters agree more frequently than random, Gwet's agreement coefficient was also calculated. The results indicate significant agreement for both Conceptualization (AC1 = 0.823, p < .001) and Permissibility (AC2 = 0.801, p < .001).

Variable	Coefficient	Value	Std. Error	95% CI	p-value
Conceptualization	Gwet's AC1	0.823	0.029	(0.765, 0.881)	< 0.001
Permissibility	Gwet's AC2	0.801	0.020	(0.761, 0.840)	< 0.001

Table 2: Gwet's agreement coefficients for inter-rater reliability.

Gwet's Agreement Coefficients 1 (categorical) and 2 (ordinal) bypass the complications mentioned in Table 1, above, by calculating the chance of expected agreement differently. For conceptualization (AC1 = 0.823; p < 0.001), the raters agreed strongly and significantly, with a 95% confidence of agreement (CI 0.765, 0.881) between good and very good. For permissibility (AC2 = 0.801; < 0.001), the raters also agreed strongly and significantly, with a 95% confidence of agreement (0.761, 0.840) between good to very good. Therefore, Table 2 shows that, in contrast to Table 1, the rating system was reliable.

3.2.3 Rater and Institutional Differences

To evaluate the nature of the differences in scores between raters, standard deviations (SD) were also calculated. The results reveal that Rater 3 scored higher on average, rater 2 lower than average, and rater 1 in-between. Variations in scores between institutions (SD = 0.682) is double that of the rater variance (SD = 0.323), indicating significant differences between institutions more so than between raters.

Rater	N	Mean	Std. Dev.	Std. Error	95% Conf. Interval
1	146	5.25	1.05	0.09	5.08, 5.43
2	146	4.88	1.25	0.10	4.68, 5.09

Rater	N	Mean	Std. Dev.	Std. Error	95% Conf. Interval
3	146	5.53	0.95	0.08	5.37, 5.68

Table 3: Permissibility statistics by rater.

Rater Pair	Mean Difference	95% Conf. Interval
1 vs 2	0.37	0.20, 0.54
1 vs 3	−0.27	−0.42, −0.13
2 vs 3	−0.64	−0.85, −0.44

Table 4: Permissibility mean score differences between rater pairs.

Between	SD
Raters	0.323
Institutions	0.682

Table 5: Permissibility standard deviations by raters and institutions.

Rater 2 (mean = 4.88) rated permissibility lower on average than Rater 1 (mean = 5.25) who, in turn, was lower than Rater 3 (mean = 5.53), who gave higher permissibility ratings. We also see in Table 3 that Rater 3 had the smallest SD (0.95), showing consistent categorical rating. Rater 2, in contrast, had a larger SD (1.25), showing a greater spread of the ratings available. The spread between raters 2 and 3 confirms this finding in Table 4. While the difference is real, it is one of degree not category (i.e., ratings may differ but remain within the assigned categorical range), as we see in concert with Table 2.

With an average SD within the institutional permissibility ratings at 0.682 and the SD between rater scores at 0.323, the disagreement between ratings is more than double than between raters in Table 5. Rater disagreement, then, is less of a complication than the ratings between institutions. In other words, while the raters differed, the way institutions

differ is greater. Thus, raters vary slightly but institutions vary significantly, which we see confirmed in the figures.

3.2.4 Accounting for Random Effects

To examine whether permissibility covaries with conceptualization while accounting for institutional and rater variance identified above, the Cumulative Link Mixed Model (CLMM) takes permissibility as an ordinal outcome and conceptualization as a predictor with random effects factored for policies (universities) and raters. The tables demonstrate that, when compared to the TOOL baseline, combination (Estimate = 3.91, $p < .001$) and RESOURCE (Estimate = 4.75, $p < .001$) were significantly associated with higher permissibility potential.

Predictor	Estimate	Std. Error	z value	p-value
Conceptualization-Combination	3.91	0.45	8.62	< 0.001
Conceptualization-Resource	4.75	1.37	3.47	< 0.001

Table 6: Fixed effect estimates to predict policy permissibility.

Random Effects	Variance	Std. Dev.
Institution	3.16	1.78
Rater	1.28	1.13

Table 7: Variance components for the random effects of both institution and rater.

Considering institutional variance (3.16) and SD (1.78), permissibility across institutions differs substantially. One might think this likely the case without rater scores. The model accounts for this variance, though, which suggests that despite the rating differences, some universities are more restrictive or permissive than others, which we see in the figures. Considering rater variance (1.28) and SD (1.13), ratings differed substantially, too, which has been explained. The model nevertheless adjusted for the

rater bias by using it as a random effect, which suggests that the central finding was not imperiled by rater disagreement. By treating institutional and rater variances as random effects, the model shows the fixed effects in Table 6 clearly.

4 Discussion

This study examines whether a correlation between how universities conceptualize AI and their permissibility could be found, which occasions two questions: (1) Do universities that call AI a TOOL attach to more restrictive policies? and, (2) Do universities that call AI a RESOURCE attach to more permissive ones? Institutional ratings and densities reveal distinct and significant clusters, voids, and correlations that answer these questions no and yes, respectively, in Figure 2 and Figure 3 for (1) and Figure 3 and Table 6 for (2).

4.1 Implications

4.1.1 The Gradation of Tools (1)

In Figure 2, given that the raters evaluated universities on two axes, conceptualization and permissibility, visualizing the same tests the assumption that both measures covary. Stratification among the permissibility states stems from conceptualization as discreet data. At the lowest stratum for TOOL, the permissibility gradient is complete. As the strata progress, TOOL naturally gives way to combination and RESOURCE, where restriction, in turn, is supplanted largely then completely by the balanced and the permissive.

In Figure 3, while many of the universities that conceptualize AI as a TOOL are restrictive (16.44%), the majority remain in balanced (53.42%) and permissive (17.81%). Therefore, insofar as TOOL can attach to restrictiveness, it is not intrinsically or inherently so. In other words, TOOLS are used both permissively and non-permissively. This is likely due to the fact that we call many technologies "tools" and that, when benchmarking policies, tool becomes something that is mirrored in the language used by peers at peer institutions.

4.1.2 The Clarity of Resources (2)

While few universities conceptualize AI as a RESOURCE in Figure 3, all are permissive (1.37%) and a balanced mix (0.68%). The absence of a balanced resource, restrictive mix, and restrictive resource indicates an extraordinary and telling void: resource does not lead to restriction. These findings indicate that for universities that adopt a resource-forward policy framework, RESOURCE acts as a hedge against restrictiveness.

This result was tested in Table 6 and Table 7, where conceptualization of AI can be seen as a statistically significant predictor of the permissibility of the AI policy. Both combination (3.91) and resource (4.75) carry higher likelihoods of more permissive policies than tool qua baseline. Universities that tend to conceptualize AI as a combination have approximately 50 ($e^{3.91}$) times higher odds of having a more permissive AI policy than tool. With resource, the odds are approximately 116 ($e^{4.75}$) times higher. The relationship is significant: at nearly 116 times higher odds of resource attaching to the permissive more so than even the nearly 50 times higher odds a combination does, it is clear that resource conceptually covaries with permissibility.

4.2 Limitations

The leading limitation of this study was self-imposed as a function of method, namely that, to avoid pressure bias, there could be no re-calibration or adjudication. The subjectivity of the raters (a professor and 2 undergraduate research assistants) led to significant, subjective differences (of interpretation) in the ratings. While our use of a mixed-methods model overcame this concern, future research could adopt different methods (e.g., non-hierarchal training sessions) to nevertheless improve inter-rater reliability further.

5 Conclusion

This study captures AI policies in time. When the raters revisited websites just one term later over the same academic year (2024–2025), many policies had changed. Many will change again. When factoring in the

steady march of AI proliferation, the "integration of AI technology into existing platforms has rendered these frameworks obsolete," thereby necessitating policies "which will likely have to be rewritten again and again" (Janos 2024).

At the same time, AI programs, platforms, and services are changing. Better still, language changes over time. Future research in this space should include mapping shifts in language as AI advances and policies respond thereto. For the moment, RESOURCE does not lead to restrictiveness and TOOL is too diffuse in use, meaning, and likely mimicry to attach meaningfully to either restrictive or permissive policies.

We no longer call calculators tools or the Internet a resource. Rather, descriptions have faded into simplicity, and these things are called now by their names. Just as with other former tool then resources heretofore, LLMs could become known as LLMs, categorically, in the future. Whether the metaphorical distinction of LLMs as something else fades could depend upon whether they are made widely available and fully integrated into education by name. To the extent that LLMs are being made more accessible, generally, and deployed by universities, particularly, the possibility that they might integrate by name into constellations of mode, method, and process is possible. Thus, while the concepts used to contain LLMs now have two paths with differing attachments to permissibility, RESOURCE carries the highest possibility, in time, of erasure of distinction. For now, though, RESOURCE carries the the highest possibility of permissibility.

Publication Details and Disclosures

Editorial Note

Since this research was supervised by the managing editor of the journal, the peer-review process was overseen by the associate editor.

Acknowledgments

The authors would like to thank the peer reviewers for their thoughtful feedback.

Funding

No funding supported this research.

Generative AI Use

The authors consulted LLMs for code generation capable of handling the plots, related calculations, and strategic recommendations.

Biographies

Miller is an assistant instructional professor in the University Writing Program at the University of Florida, where he serves as the AI strategist and teaches professional writing courses.

Sachdeva is a fourth-year undergraduate student majoring in Biology at the University of Florida.

Walker is a junior at the University of Florida. He is currently pursuing a Bachelor of Science Business Administration Information Systems degree in addition to a certificate in Artificial Intelligence.

References

DELEUZE, Gilles and GUATTARI, Félix, 1994. *What Is Philosophy?*. New York: Columbia University Press.

HILL, Kashmir, 2025. The Professors Are Using ChatGPT, and Some Students Aren't Happy About It. *The New York Times*. Online. 14 May 2025. Available from: https://www.nytimes.com/2025/05/14/technology/chatgpt-college-professors.html

JACOBS, Keith and MANZI, Tony, 1996. Discourse and Policy Change: The Significance of Language for Housing Research. *Housing Studies*. 1996. Vol. 11, no. 4, , p. 543–560. DOI 10.1080/02673039608720874.

JANOS, Nik, Zach Justus, 2024. Your AI Policy Is Already Obsolete. *Inside Higher Ed*. Online. 22 October 2024. Available from: https://www.insidehighered.com/opinion/views/2024/10/22/your-ai-policy-already-obsolete-opinion

KOCHIS, Bruce, 2005. On lenses and filters: The role of metaphor in policy theory. *Administrative Theory & Praxis*. 2005. Vol. 27, no. 1, , p. 24–50.

LEGATT, Aviva, 2025. Federal AI Policy Is Here. Are Universities And Schools Ready?. *Forbes*. Online. 2 May 2025. Available from: https://www.forbes.com/sites/avivalegatt/2025/05/02/federal-ai-policy-is-here-are-universities-and-schools-ready/

MENZ, Mareike M., BLANGERO, Annabelle, KUNZE, Damaris and BINKOFSKI, Ferdinand, 2010. Got it! Understanding the concept of a tool. *NeuroImage*. Online. 2010. Vol. 51, no. 4, , p. 1438–1444. DOI 10.1016/j.neuroimage.2010.03.050.

MURRAY-RUST, Dave, NICENBOIM, Iohanna and LOCKTON, Dan, 2022. Metaphors for designers working with AI. In: *DRS2022: Bilbao*. Online. Bilbao: Design Research Society. June 2022. p. 1–19. DOI 10.21606/drs.2022.667.

ROBERT, Jenay and MCCORMACK, Mark, 2025. *2025 EDUCAUSE AI Landscape Study: Into the Digital AI Divide*. Online. Boulder, CO. Available from: https://www.educause.edu/content/2025/2025-educause-ai-landscape-study/

SINGER, Natasha, 2025. Welcome to Campus. Here's Your ChatGPT. *The New York Times*. Online. 7 June 2025. Available from: https://www.nytimes.com/2025/06/07/technology/chatgpt-openai-colleges.html