

TABLE OF CONTENT

Anita Christina Anthony Pre-Service ESL Teachers' Mindsets and their Approaches towards Assessments https://doi.org/10.24191/mvahvg19	1–21
Maryam Muidh Alsulami, Sharifah Shahnaz Syed Husain, Aini Faridah Azizul Hassan Saudi EFL Learners' Attitudes Towards Arabic-English Code-Switching https://doi.org/10.24191/yjvd8a60	22–37
Julazzmie Kambutong, Nur Farha Shaafi Content Validity of a Survey on Knowledge, Skills, and Readiness for Generative AI https://doi.org/10.24191/xnvcmw23	38–57
Aini Syahira Jamaluddin, Nabila Huda Nasir, Nur Azyyati Mohamad Sa'edin, Ummu Habibah Mohd Sakri @ Shukri Can Artificial Intelligence Replace Teachers? Perspectives from Pre-Service Teachers in Elementary Education https://doi.org/10.24191/vjk6sz19	58–73
Xia Yutong, Talaibek Musaev Grammar of Obligation: Syntactic Patterns in Malaysian Reddit AIED Discourse https://doi.org/10.24191/rs3g1v95	74–97
Mohd Hafiz Mat Adam, Geethanjali Narayanan, Juriani Jamaludin, Rafidah Abd Karim Fostering Independent Expository Essay Writing Development Using Song-Assisted Learning (SAL) to Avert AI Over-Reliance https://doi.org/10.24191/nhh8d038	98–112
Farhana Atiqah Ahmad Radzuan, Marina Mohd Arif Teacher Perspectives on ChatGPT Integrated Vocabulary Instruction in a Flipped ESL Classroom https://doi.org/10.24191/9deysg59	113–135
Damia Syafiqah Mohd Salleh, Norlina Mohd Sabri, Fazlin Marini Hussain SpeakEasy Nihongo: Mobile Application for Japanese Beginner Learners https://doi.org/10.24191/1k11n645	136–152



Grammar of Obligation: Syntactic Patterns in Malaysian Reddit AIED Discourse

Xia Yutong

s2188193@siswa.um.edu.my

Department of Asian and European Languages

Faculty of Languages and Linguistics

Universiti Malaya, Malaysia

Talaibek Musaev*

musaev@um.edu.my

Department of Asian and European Languages

Faculty of Languages and Linguistics

Universiti Malaya, Malaysia

Corresponding author*

Received: 6 January 2026

Accepted: 24 February 2026

Published: 25 May 2026

CITE THIS ARTICLE:

Xia, Y., & Musaev, T. (2026). Grammar of obligation: Syntactic patterns in Malaysian Reddit AIED discourse. *Journal of Creative Practices in Language Learning and Teaching*, 14(1), 74–97. <https://doi.org/10.24191/rs3g1v95>

ABSTRACT

This paper examines how Malaysian Reddit discourse constructs the roles of the state, market, and individual in AI-facilitated education. Using Appraisal analysis alongside topic modelling (LDA and BERTopic), the study analyses public posts from r/Malaysia spanning 2024 to 2025, identifying three overlapping discourse frames: fiscal governance, corporate innovation, and individual obligation. These frames are interwoven to produce what we call a grammar of obligation, in which the modernization of education is described as a distribution of responsibilities across distinct social actors. The state is cast as a limited fiscal provider, corporations as technological enablers, and students as agents of constant self-adjustment. This tripartite arrangement encodes precarity and systematically shifts accountability onto individuals, framing employability as a personal outcome rather than a shared policy responsibility. Combining computational modelling with discourse semantics, the paper demonstrates how macro-level economic structures shape everyday public debate around fairness, competence, and value in the Malaysian context. A central theoretical finding is temporal misalignment across the three frames, whereby policy operates in the iterative present, technology in the progressive aspect, and individual obligation in the deontic future, normalizing coordination failure as personal



navigational deficit. The findings illuminate how Reddit's participatory format reinforces neoliberal logic through registers of fiscal realism, pragmatic optimism, and credential anxiety.

Keywords: AI Education; Discourse Analysis; Systemic Functional Linguistics; Syntactic Construction

INTRODUCTION

Following the release of ChatGPT in late 2022, AI-driven Educational Technology has improved its pace between pilot projects to agenda-setting infrastructure in education systems all over the globe (DeVito et al., 2025). Once separate tools, namely personalized feedback engines, intelligent tutoring systems, and automated assessments, are now integrated into broader reform programmes, with AIED increasingly positioned to restructure pedagogy, redistribute instructional labour, and reshape educational governance (Alfredo et al., 2024; Schleicher, 2024). The national digitalization plans that anticipate equity and inclusion as organizing principles are the way in which this international trend is being served in Malaysia. Policy initiatives approach AIED as an instrument not only to support productivity but also to create a mechanism of structural disparity in access, quality, and results, particularly within linguistic, regional, and socioeconomic categories (Jamaluddin et al., 2025). A recent systematic review of AI integration in language education similarly finds that while AI tools expand learning access, their implementation raises persistent concerns around equity, ethical transparency, and institutional readiness, particularly in under-resourced contexts (Ibrahim Brian, 2025).

Under this horizon, AI literacy comes into play as an essential competency on top of the functional knowledge of data and algorithms. It involves the practical and moral ability to question outputs of models, discuss feedbacks and use the interactions to generate self-directed learning and societal involvement (Lin et al., 2025). The recent governmental discourses of inclusive and culturally based AI education make the technology subordinate to a broader national initiative of inclusive digital transformation, placing the technology in the local, pluralistic context of linguistic and cultural Malaysia (Ministry of Education, 2023; Bernama, 2024).

To achieve these aspirations, analytic lenses have to look out of the procurement and implementation metrics. The adoption of AIED is saturated by discourse: policies, platform posts, official statements, expert commentary, and everyday talk that render technology legible, desirable, risky, or obligatory (Mansoor et al., 2024). These discursive processes distribute agency among ministries, educators, ed-tech vendors, civil society, parents, and students; they confer credibility and responsibility; and they construct temporal horizons that present reform as immediate, pending, or contested (Lewis, 2025; Viberg et al., 2024). Accordingly, the present study shifts the evidentiary centre of gravity from outputs and outcomes to the linguistic practices through which AIED becomes socially meaningful in Malaysia. It asks not only what Malaysians say about AI in education, but how they say it, which grammatical resources they recruit to legitimate or resist AIED, and how these resources are shaped by the communicative architectures through which public discourse circulates.



Reddit provides a theoretically instructive site for this inquiry. Unlike broadcast-oriented platforms, Reddit is organized through community segmentation into subreddits with explicit rules, volunteer moderation, and rank-ordered visibility governed by up- and down-votes (Stevens, 2021). In Malaysia, r/Malaysia and adjacent communities function as arenas in which policy news, personal experience, expert explanation, and critical debate intersect (Kamaruddin et al., 2024). The result of these affordances is a discursive field, which is decentralized, patterned and contingent on visibility, temporally sensitive and negotiated within the norms of the local communities, but whose frames of interpretation circulate unequally across the subcultures (Stevens, 2021; Kamaruddin et al., 2024). This type of structure is analytically useful in that it provides the ability to trace the manner in which the logics of platforms such as voting, moderation, as well as segmentation affects the circulation of language regarding authority, accountability, expertise, and public interest.

On this foundation, Critical Discourse Analysis provides the overarching orientation to language as social action (Fairclough, 2005), while Systemic Functional Linguistics (Halliday, 1978) specifies how grammatical choices realize social meanings. Transitivity configurations are examined to determine how agency is allocated among institutional and lay actors; voice and modality are analysed to reveal how obligation, certainty, and evidential stance are stratified; and tense- aspect- mood distributions are traced to model temporal horizons and reform rhythms. The Appraisal framework (Martin & White, 2005; Dai & Pan, 2024) operationalizes evaluative language, including attitude, engagement, and graduation. This allows the study to map how trust, risk, benefit, and harm are discursively calibrated. This approach to modality analysis has been similarly applied in recent Malaysian media discourse research (Suppiah et al., 2025), supporting its applicability to digitally mediated public texts.

To interpret temporal and affective undercurrents that exceed proposition-level semantics, the analysis mobilizes Koselleck's distinction between the "space of experience" and the "horizon of expectation" , hereby making visible how legacies of pedagogy and policy inflect present evaluations of AIED, and how pre-articulated collective emotions (anticipation, fatigue, optimism, ambivalence) circulate beneath explicit claims, shaping receptivity before positions congeal (Koselleck, 2004; Robinson, 2025).

As education systems worldwide grapple with AIED adoption, the debate is not merely technical but normative, centring on who bears responsibility for outcomes. In neoliberal governance models, the state retreats to regulatory and fiscal functions, markets are cast as engines of innovation, and individuals are constructed as self-optimising agents navigating perpetual uncertainty (Chiu et al., 2024; Fairclough, 2005). While this framework provides analytical traction, it is worth noting that distributed accountability may also reflect deliberate policy pluralism or collaborative governance logics not reducible to neoliberalism alone (see e.g., Jessop, 2002; Moe, 2015).

This paper will put forward the thesis that the discourse on AIED at Malaysian Reddit builds such a fragmented accountability order, a grammar of obligation, through linguistic construction. The stakes are both linguistic and political. By showing how transitivity patterns, modal configurations, and temporal encodings assemble this distributed-obligation system, we make visible the mechanisms through which AIED discourse normalises precarity as a personal navigational



problem rather than a governance failure. This matters because discourse is productive, not merely descriptive: it shapes what can be thought, done, and held to account in education policy (Fairclough, 2015). If Malaysian publics can articulate AIED only through grammars that fragment responsibility, reform is linguistically channelled toward individual adaptation strategies instead of systemic coordination.

Hence, this study's central questions are therefore both linguistic and infrastructural:

- 1). What dominant discourse frames structure Malaysian Reddit discussions of AIED (2024-2025)?
- 2). How do these frames grammatically distribute agency and obligation across state, market, and individual actors?
- 3). What does the co-occurrence of syntactic resources reveal about the linguistic naturalization of fragmented accountability?

By correlating visibility dynamics with patterns in transitivity, modality, and Appraisal, and by reading these patterns through temporal and affective lenses, the analysis specifies the mechanisms by which platform architectures refract national aspirations for equity-oriented AI education. In doing so, it provides an empirically grounded account of how Malaysian public discourse constructs the possibilities and limits of AIED, clarifying when and how these technologies are made to democratize participation, consolidate existing hierarchies, or do both at once across differently positioned communities.

METHODOLOGY

Data Collection and Preprocessing

This study scraped Malaysian Reddit communities, principally r/Malaysia. This study adopts open-source code (Stevens, 2021) with Python requests and Selenium. The window spans 2024–June 2025, a period of rapid LLM diffusion across sectors and heightened debate on AI's social impacts. Malaysia's push for inclusive digital transformation provides the context for examining how AI-enabled educational technology (AIED) is publicly constructed.

Malaysia-specific queries (e.g., "AIED", "AI in education", "AI-generated content in learning") yielded 975 posts. After deduplication, language filtering, and relevance/engagement screening (a threshold of at least 5 upvotes and at least 3 comments) against policy, pedagogy, administration, and equity criteria, 314 items (posts with associated comments) remained for topic and stance analyses. It should be noted that this threshold prioritises high-visibility exchanges, which may over-represent the views of engaged, urban, and digitally active users and thus may not fully capture the range of perspectives present in broader Malaysian public discourse.

For finer analysis, we applied a standard SFL procedure to segment the 314 posts into clause-level units expressing single propositions, yielding 598 clauses (≈ 1.9 per post). Reddit AIED posts often braid multiple frames in one text (e.g., criticizing budgets, endorsing corporate innovation, and worrying about credential inflation), so clause-level annotation lets each meaning unit be assigned to its proper frame. Frame assignment followed a hybrid procedure: topic model outputs (LDA



and BERTopic) provided an initial thematic label for each clause, which was then reviewed by the lead researcher. To mitigate subjective bias and refine the coding scheme, a second researcher reviewed a 15% subset of the data. Disagreements were resolved through discussion to reach a consensus (negotiated agreement), rather than through blinded independent coding. Final frame assignments were then completed by the lead researcher, with intra-rater reliability confirmed via a test-retest procedure on 15% of the data (91.4% agreement) and a boundary-case log; conservative inclusion minimized false negatives. All frame-distribution statistics in §3 therefore use the 598-clause set as the denominator, not the 314 posts.

Preprocessing addressed Malaysia's multilingual, code-switching ecology. High-frequency Malay function words (e.g., *yang*, *dengan*, *untuk*, *boleh*, *akan*) were mapped to English to stabilize syntax, while culturally loaded lexemes (e.g., *kampung*, *rakyat*, *kacang lupukan kulit*) were retained. The English stream underwent NLTK stop-word removal and spaCy lemmatization; Malaysia-specific proper nouns and multiword terms (e.g., *Kuala Lumpur*, *smart_classroom*) were preserved as fixed tokens. This normalization procedure may have partially affected the transitivity and modality analysis, since Malay modal particles such as *boleh* (can/may) and *akan* (will) carry distinct aspectual and deontic nuances that do not map precisely onto their English equivalents. Where such mappings were applied, the resulting syntactic patterns should be interpreted with this limitation in mind.

Ethics procedure followed AoIR (2020): only public content was used; platform ToS and robots.txt were respected; identifiers were anonymized; data are research-only and will be destroyed five years post-project.

However, this research admits to several limitations, including elite-discourse bias resulting from engagement thresholds, substantial sample contraction (975 → 314, -67.8%), and algorithmic visibility effects; findings thus represent dominant, high-engagement patterns rather than population sentiment. In particular, the engagement threshold systematically excludes low-visibility posts that may reflect the views of less digitally active or less urban communities, meaning the corpus is best understood as a window into dominant public discourse rather than a representative cross-section of Malaysian opinion on AIED.

What is more, a key methodological compromise was the inevitable process of mapping high-frequency Malay function words to English equivalents. We acknowledge that this procedure simplifies the specific grammatical configurations to a fine-grained SFL analysis.

Model Performance Analysis

This study adapts analytical procedures from Liu et al. (2024), Verbytska (2024), and Zhang et al. (2024). Cleaned Malaysian Reddit texts were modelled with Latent Dirichlet Allocation (LDA) and BERTopic. LDA (Blei et al., 2003), implemented via spaCy and Gensim, treats each document as a mixture of latent topics, each defined by a word distribution, enabling multi-topic assignment and transparent word–topic probabilities. BERTopic (Grootendorst, 2024) complements this with BERT embeddings (Devlin et al., 2019) and class-based TF-IDF to form dense, semantically coherent clusters that typically yield a single dominant topic per document. Embeddings capture contextual similarity among lexemes and thus separate adjacent themes in code-switched,



multilingual discourse. Following Verbytska (2024), both models were pre-tuned and jointly interpreted to balance probabilistic breadth with embedding-based specificity.

Experimental evaluation prioritized topic coherence (C_v) for semantic validity and used perplexity as a secondary fit diagnostic (Mimno et al., 2011). Optimizing topic number produced a peak C_v of 0.5524 for LDA at $k = 5$, indicating an effective balance between granularity and cohesion on this corpus (Figure 1). A C_v score in the range of 0.5 to 0.6 is generally considered moderate to good performance for topic models applied to social media corpora, suggesting that the identified topics are semantically coherent and meaningfully interpretable. Perplexity, computed after Brzustewicz and Singh (2021), reached 157.45, consistent with an adequate representation of latent structure. Lower perplexity values indicate better generalization to unseen data; the obtained value falls within an acceptable range for a corpus of this size and domain. Together these results indicate that LDA furnishes a stable document–topic scaffold for Malaysian Reddit discourse, particularly where subthemes are separable and vocabularies are not overly compressed.

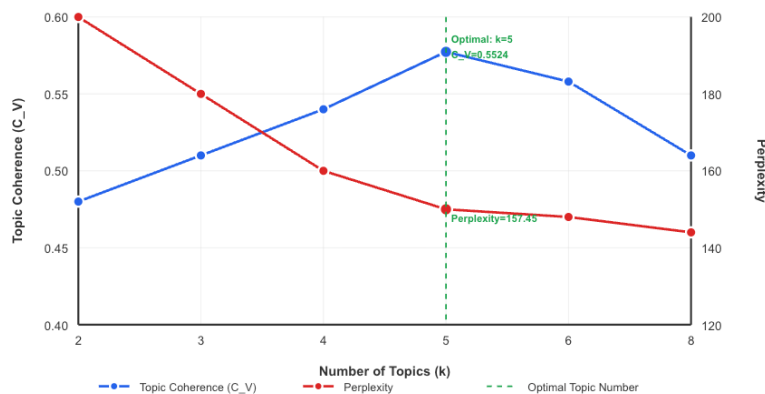


Figure 1. LDA model evaluation on Malaysian Reddit Corpus

BERTopic attained its best coherence at $C_v = 0.4651$ with $k = 4$. Although below LDA’s peak, the embedding approach proved sensitive to fine-grained lexical neighborhoods and emergent subdomains associated with code-switching, idiomatic Malay, and locally inflected policy talk. A t-SNE projection (Figure 2) shows a dispersed topology with a large, diffuse hub and several smaller, well-separated clusters, consistent with Reddit’s dialogic, community-governed dynamics.

In practical terms, LDA provides the primary thematic scaffold for qualitative appraisal analysis, while BERTopic is layered to probe intra-topic variation, identify semantically proximate neighborhoods, and flag outliers for close reading. Topics across the two models were aligned by comparing their top keywords and examining the degree of document overlap between LDA topic clusters and BERTopic clusters. Where the two models assigned high proportions of the same documents to thematically similar clusters, those clusters were treated as converging evidence for a single frame. The three final frames (Policy and Governance, Technological Innovation and Collaboration, and Higher Education Empowerment) were derived by grouping aligned topic clusters according to shared functional and evaluative orientation, as confirmed through iterative close reading of representative clauses from each group. This integrated strategy preserves



interpretability while remaining sensitive to Malaysia's heterogeneous, event-driven discourse on AIED.

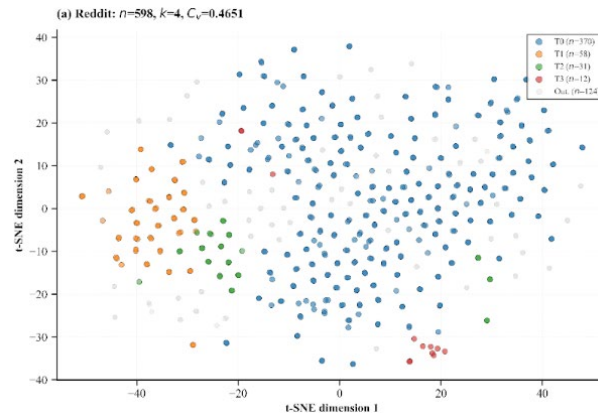


Figure 2. BERTopic Document Distribution (t-SNE Projection)

RESULTS

Building on the LDA and BERTopic results, this section advances a unified framework that links computational outputs with Martin and White's (2005) Appraisal system to analyze how Malaysian AIED discourse realizes meaning and evaluation on Reddit. Latent themes from topic modelling are mapped to Appraisal subsystems, and it aims to recontextualize distributional clusters within a functional-semantic account.

The analysis proceeds from functional representations of practice (e.g., classroom deployment, assessment, resource allocation) to interpersonal stance work (affective positioning, trust, evaluation), aligning with ideational, textual, and interpersonal meta-functions while remaining platform-specific rather than institution-centric. Operationally, we first identify how actions and procedures are represented and rationalized; next, how knowledge, expectations, and evidential commitment are signalled through engagement resources and modality; and finally, how emotions such as optimism, scepticism, anxiety, are made salient and scaled via graduation.

Within this architecture, attitude captures distributions of affect, judgment, and appreciation toward technologies, policies, and actors; engagement models dialogic positioning via attribution, projection, concession, and the management of alternative voices; and graduation tracks intensification/attenuation, a salient mechanism in a code-switching environment where Malay and English lexemes calibrate emphasis and nuance. Together, these components bridge distributional patterns with theory-informed interpretations of value, alignment, and affect. The resulting framework is organized as follows:

Frame 1: Policy and Governance — National budget allocation, institutional regulations, and curriculum standards; embodying the governance logic underpinning the institutionalization of AIED.

Frame 2: Technological Innovation and Collaboration — AI-driven breakthroughs, digital platform development, and international partnerships integrating technology and the education system.



Frame 3: Higher Education Empowerment and Industry-Oriented Collaboration Paths — Emphasizes universities as intermediary hubs in industry-education synergy mechanisms, connecting the education system with industrial demand.

These frames are also reflecting the thematic areas of the computational models of Malaysian Reddit, as well as the normative dimensions by which AIED is appositively invoked in the discussion of the topic in the public. Making the topic modeling analysis and the appraisal analysis form a dual perspective where, the first step in the analysis involves uncovering the latent thematic structure by the models, whereas the second step of the analysis involves displaying how those styles are judged, argued or scaled up or down as the users position themselves towards certain claims.

The unified framework proceeds hierarchically from institutional and functional policy orientations, implementation practices, and technological articulation. Subjective and interpersonal strata include affect, trust, and personal evaluation. This ordering aligns with the ideational, textual, and interpersonal meta-functions of Systemic Functional Linguistics (Halliday, 1978) and allows discourse to be read simultaneously as action, as information flow, and as stance-taking (Martin & White, 2005). In practice, institutional positioning and expert attribution are examined alongside representations of classroom deployment and resource allocation. These elements are then interpreted with respect to engagement resources and modality, and finally read through graduation to recover shifts in emphasis and affective load. The result is a coherent account of how Malaysian publics align with or resist AIED through interlocking narratives of policy, practice, understanding, and feeling.

On this basis, the analysis moves to corpus-level detail. Following NER identification of key organizations, ministries, universities, locations, and events, concordance analysis is conducted using Wordless (Ye, 2024) to examine co-occurrence patterns surrounding these named entities. This integrated approach, which combines NER-identified terms with concordance analysis, shows how specific linguistic constructions operate within AIED discourse. Figure 3 illustrates the procedure by treating a search term as the node and displaying its left and right distributions. The extracted entities serve as cues for tracing judgment and appreciation like competence, fairness, credibility, and perceived impact within the Appraisal system. Discourse segments linked to each thematic frame are then aggregated into subsets to support within-frame appraisal profiling and to compare stance patterns across domains such as admissions and finance, program choice, and employment outcomes.

Left	Node	Right
however hesitant commit specialized field early potentially regret later wonder	cybersecurity	major apu offer well opportunity cs truly worth investment advice
semester suppose pick major either software development internet thing iot	cybersecurity	know may suit subreddit want understand current industry demand malaysia
major would suit current demand malaysia also hear increase demand	cybersecurity	major really fancy idea deal network server take one core
want know could give decide plan perfect still possible enter	cybersecurity	without degree hey everyone want ask something bug year old
recently feel bit behind life career finally decide start learn	cybersecurity	enrol google cybersecurity professional certificate coursera thing degree Computer_Science sure
behind life career finally decide start learn cybersecurity enrol google	cybersecurity	professional certificate coursera thing degree Computer_Science sure right path follow
unsure like employer even consider someone without tech degree google	cybersecurity	cert enough land job like soc support cyber analyst focus
research cs field senior pls help major bachelor get job	cybersecurity	ai datum science really sure branch cs part ahaha also
armed force home million upgrade prison staff quarter million enhance	cybersecurity	government set national scam response centre health million subsidy mammogram
move aus canada last year master respectable uni foreign partner	cybersecurity	field malaysia uni student pursue Computer_Science degree malaysia interested diving
field malaysia uni student pursue Computer_Science degree malaysia interested diving	cybersecurity	field however base people know limited knowledge cybersecurity field sound

Figure 3. Example of concordance analysis with a query export.



As AIED evaluations on Reddit are often couched in modality, evidentiality, and technical phrasing rather than overt sentiment words, generic polarity models trained on product or movie reviews are insufficiently sensitive. To address this, a transformer model (DistilBERT-base-uncased) was fine-tuned on a stratified (Sanh et al., 2019), manually annotated sample from the Malaysian Reddit corpus (n=285) using a three-way scheme (positive/neutral/negative). Inter-annotator reliability was substantial ($\kappa = 0.71$), yielding a consensus set of 257 items for training and validation.

Given these characteristics, sentiment outputs are treated as descriptive indicators of affective tendency within Malaysian Reddit rather than as stand-alone evidence. Analytical weight rests with qualitative appraisal analysis, which recovers how stance is assembled linguistically and how intensity is modulated across communities and threads. In this design, the topic models supply the thematic scaffold, the NER layer situates evaluation in an institutional landscape, the sentiment classifier offers a coarse orientation to affect, and the appraisal framework provides the theory-driven account of how Malaysian discourse organizes values, aligns or disaligns with claims, and scales commitment and emotion in constructing AIED’s social meaning.

Frame 1: Policy & Governance - Distributed Accountability Structure

Frame 1 incorporated LDA’s Topics 1 and 5 and BERTopic’s Topics 1 and 0. It ranked second across both platforms, accounting for 33.9% (n = 203/598) of the Reddit corpus. It is structured through fiscal-procedural lexis such as “budget”, “allocation”, and “funding,” which enact political legitimacy via resource distribution rather than institutional authority. Figure 4 confirms this emphasis: accountability/transparency (35%), equity-oriented distribution (30%), monitoring (20%), and participation (15%) configure governance as transparency-driven, with surveillance and claims-making subordinate to demands for quantified allocation.

Table 1. The distribution of top-30 most salient keywords for Frame 1

LDA	BERTopic
Topic 1: million, billion, malaysia , source, free, flood, contact, twitter, aid, government, budget, allocation, funding , support	Topic 1: million, billion, malaysia , source, free, flood, contact, twitter, aid, government, budget, allocation, funding , support
Topic 5: development, investment, expenditure, financial, economic, national, ministry, public, policy, initiative, programme, resource, infrastructure, service, welfare , reform	Topic 0: development, investment, expenditure, financial, economic, national, ministry, public, policy, initiative, programme, resource, infrastructure, service, welfare , reform

Pattern 1: [Named Leader] + [Fiscal Verb] + [Quantified Resource] + [Beneficiary Category]

- Prime Minister Anwar Ibrahim madani government third budget education ministry 10 billion
- Government introduce special allocation RM500 million for AI literacy development



Pattern 2: [Budget/Allocation] + exceeds/marks/impacts + [Quantified Scale/Community]

- Budget allocation exceed billion first time Malaysia history.
- Budget mark significant milestone total allocation.

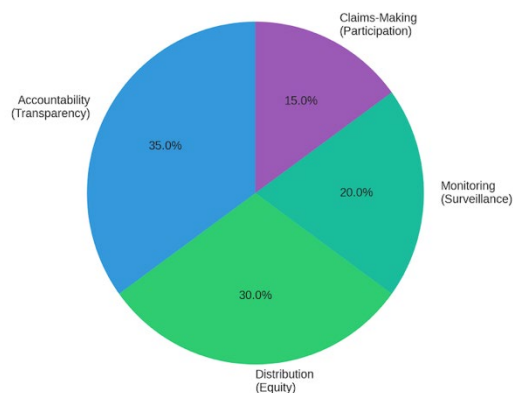


Figure 4. Functional composition in Frame 1

The corpus reveals systematic clausal structures encoding fiscal logic and multi-party accountability. NER concentrates agency in identifiable actors: PM Anwar Ibrahim (n=23), named ministers (n=18), and institutional entities such as Ministry of Education (n=47), MDEC (n=42). They all directly linked to fiscal commitments. The obligatory quantitative modifier (156 MONEY entities) constructs “calculable accountability”: governance legibility through precise figures (“third budget education ministry 10 billion”).

Ditransitive syntax dominates (Pattern 1, n=70): Government (Subject) – Fiscal Resource (Direct Object) – Beneficiary (Indirect Object). This configuration naturalizes education governance as fiscal negotiation, not pedagogical expertise. AIED emerges within budgetary items (digital infrastructure, technology grants), reframing AI education as a resource distribution problem: rural coverage, community investment equity, urban-rural gaps.

Alternative syntactic configurations (Pattern 2) shift agency to fiscal instruments themselves. Budgets “exceed” thresholds, allocations “mark” milestones, and funds “impact” sectors, temporarily displacing human actors to naturalize governance-as-measurement. Yet this objectification operates within a broader accountability framework: every allocation remains traceable to designated decision-makers, constructing budgets as political products subject to citizen audit rather than self-executing technical processes.

Budget cycles are time-encoded iteratively: the funds are announced, renewed and updated. Therefore, policy follows periodic repetitions as opposed to a linear curve, with immediacy supported by monetary pledges. Governance is narrated mainly in the simple present (e.g., “allocates”, “provides”, “marks”). This construes governance as ongoing practice. The present perfect (“has increased,” “has reached”) frames milestones as already achieved. It also stabilizes legitimacy claims. The past–present contrast (“was sufficient” / “is not”) compresses decline and justifies urgency. Budget-cycle futurity (“will fund”, “will expand coverage”) projects near-term deliverables. It does not close uncertainty. In totality, there is a shift in the temporal staging



between enactment of the present, perfective attainment, to a prospective rollout. It maintains policy as readable when cycling over policy cycles without limiting space to critique the policy in a cycle.

Figure 5 uncovers the sources of negativity (total: -0.96): The sources of negativity create by Named Leader Blame, Regional Inequality, Historical Comparison, and Quantified Insufficiency are strong in negative sentiment. Such an individualized structure of accountability is aimed at where identifiable decision makers are involved with the negativity being concentrated in a location where fiscal accountability allows deficits to be traced.

On Reddit, the full distribution of the fine-grained affect shows that it is mostly neutral (73.5%), with Hope (8.9%), Anger (8.3%), Disappointment (5.7%), and Anxiety (3.6%) distributed around budget talk (“education budget lower than defence”). Graduation erects itself on quantification to compound the stance, commitment on scale to compound the commitment and Engagement on named attribute to narrow the dialogic space. Taken collectively these resources constitute AIED legitimacy as procedural justice: credibility follows transparent, traceable, and constituency-based bargaining within the ethnolinguistic Malaysian environment, negativity plays the role of an accountability request and not a sign of governmental failure. Consistent with this orientation, the grammar of Frame 1 places a specific accountability emphasis: Clauses link named leaders to quantified allocations. Many are ditransitive. They foreground fiscal transparency. They present budgets as actions to justify to diverse publics.

When AIED implementation is questioned, explanations in this discourse commonly pivot to “insufficient funding,” whereas pedagogical design, technological fit, and employment coordination are more often discussed in Frames 2–3. Consequently, government is frequently described as a fiscal sponsor rather than an education architect, a portrayal that becomes salient as corporate actors (Frame 2) and individual students (Frame 3) are invoked to bridge the space between public investment and market demands.

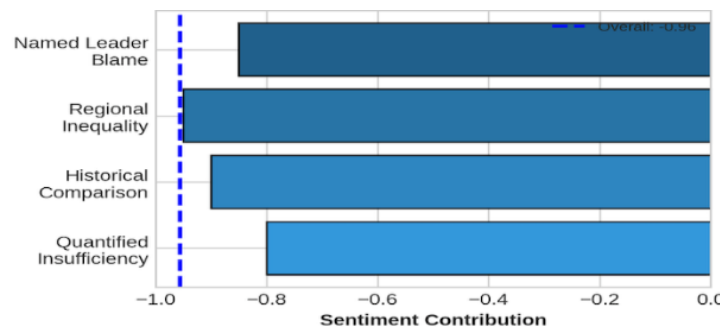


Figure 5. Linguistic negativity sources in frame 1 Reddit

Frame 2: Technological Innovation & Collaboration - Pragmatic skepticism

Frame 2, integrating LDA’s Topics 2 and 3 with BERTopic’s Topic 3, emerged as the dominant frame, encompassing 35% (n = 211/598) of the Reddit corpus. As illustrated by Figure 6, frame 2 reveals innovation discourse operating through corporate agency rather than state-led production, exhibiting cautiously positive affect (1.5:1 positive-to-negative ratio) and a fragmented functional structure that positions Malaysia as a technology adopter rather than producer.



Table 2. The distribution of top-30 most salient keywords for Frame 2

LDA	Reddit
<p>Topic 3: help, machine, model, practical, development, advanced, research, capability, solution, transform, upgrade, implement, enhance, integration</p> <p>Topic 2 : technology, innovation, platform, system, smart, application, tool, datum, service, create, improve, method, intelligent, digital, use, ,</p>	<p>Topic 3: technology, ai, innovation, korea, malaysia, communication, transcelestial, ytl, regulate, capability, artificial_intelligence, partnership, collaboration, company, enterprise, cooperation, international, development, platform, system, digital, advanced, modern, breakthrough, solution, integration, implementation, infrastructure, connectivity</p>

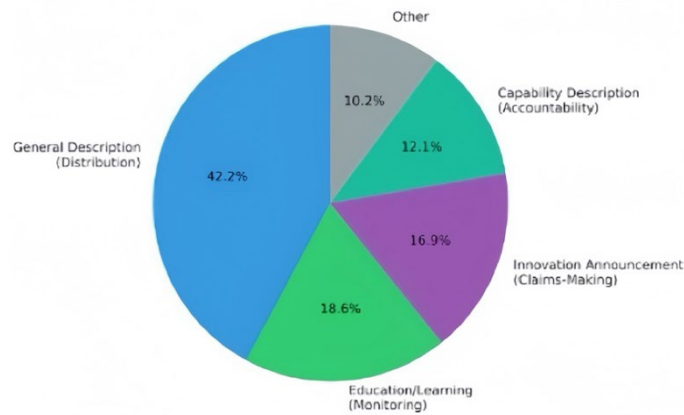


Figure 6. Comparative functional composition across platforms in Frame 2

Pattern 1: [Corporate Actor] + [Action Verb] + [Technological Initiative] + [Contextual Qualifier]

- MDEC launched AI literacy program partnering with Microsoft Azure—but implementation delayed due to budget constraints
- Google establishing AI research hub in Kuala Lumpur, focusing on Southeast Asian language models

At the clausal level, pattern 1 operates through a “corporate subject – transitive verb – technological object” configuration that foregrounds institutional agency instead of individual accountability.

A lexical inventory identifies five recurring corporate actor types in the corpus. Local telecoms (YTL, Maxis; 31 mentions) “deploy,” “expand,” and “connect,” which index infrastructural implementation. State-owned enterprises (Petronas, TNB; 24 mentions) “collaborate,” “support,” and “fund,” which balance market interests with national development roles. Government agencies (MDEC, MOSTI; 47 mentions) “launch,” “promote,” and “regulate,” which signal policy



coordination and institutional acceleration. A lexical inventory identifies five recurring corporate actor types in the corpus. Local telecoms (YTL, Maxis; 31 mentions) “deploy,” “expand,” and “connect,” which index infrastructural implementation. State-owned enterprises (Petronas, TNB; 24 mentions) “collaborate,” “support,” and “fund,” which balance market interests with national development roles. Government agencies (MDEC, MOSTI; 47 mentions) “launch,” “promote,” and “regulate,” which signal policy coordination and institutional acceleration. International technology companies (Google, Microsoft, Alibaba, 56 mentions) partner, provide, and integrate, and this confirms the need to get technology as an outside-in factor. The local startups (18 mentions) customize, adapt, and pilot, which implies the foreign technologies localization.

Besides, technological objects are made grammatical subjects, which are allowed temporarily replacing corporate actors and making innovation naturalized: a technological capacity, not an organizational decision-making. This change of grammar builds the effect of technology as an independent entity and clouds the institutional reliance and limitations of resources on implementation.

Pattern 2: [Malaysia] + [Comparative Verb] + [Regional Competitor] + [Quantified Gap] + [Domain Specification]

- Malaysian CS graduate prefer Singapore higher salary 40% more
- AI patent filings significantly lower than Korea, only 15% of their annual output

In the Pattern 2, being a comparative structure, Malaysia is placed in a latecomer technological-development environment, as other countries play different rhetorical roles. Such modes of comparisons like insufficient or lagging, are slightly negative in their connotation, and their usage creates a relative catching up logic. It implies that Malaysia’s progress unfolds along pre-established trajectories. Singapore and South Korea often serve as reference points. Incorporating with comparative prepositions (“compared with”) and verbs (“catch up”), they co-create a competitive benchmarking narrative. Hereafter, Innovation is framed as continuous assessment and self-correction instead of the autonomous capability building.

Interestingly, the Malaysia-China combination (34 mentions) operates simultaneously as an avenue of collaboration and a venue of inter-country tension with regard to technological dependency and data management. Malaysia-Thailand-Vietnam triad (29 mentions) supports the idea of regional benchmarking, whereas the Malaysia-Indonesia-Philippines cluster (12 mentions) shows the aspect of the lateral self-positioning. Combined, these comparative constructions effectively frame Malaysia as a regionally situated in an innovative hierarchy, which establishes modernization as a continuing catch-up game amid regionalized competition. Malaysia has this relational orientation displayed externally through this framing. It views innovation as a relative excellence as opposed to a self-sufficient accomplishment.

These gaps are grammaticalized by recurring quantitative markers (3:1, 40 percent, 15 percent) as factual indices that can be measured and thus made linguistically observable, which makes developmental hierarchy measurably observable. Such comparative syntax focuses discussion on concrete policy-linked domains: education, infrastructure, human capital, and R&D investment. Thereby imparting a policy-oriented tone to the innovation discourse.



Pattern 3: [Qualifying/Skeptical ADJ] + [Technology NOUN]

- Limited AI adoption
- Nascent technology ecosystem

A third tendency manifests itself in ADJ + NOUN structures which meet different discursive functions and affective orientations. Ordinary modifiers such as limited, developing, and challenges build the cautious technological advantage, but with reference to its immaturity, the issues it encounters and the constraints of a real world. These connotative modifiers tend to moderate the conviction and utopian fantasy of the technology and indicate the doubts of consumers regarding the potential of current AIED and systemic assistance. These play the role of an amplifier of capability belief as well as a buffer of cognitive alertness (Taboada, 2016).

Paired with this adjective-noun phrase structure, temporal coding follows the project cycle: It reflects in verbs predominantly using the present progressive and present perfect tenses. This pattern constructs innovation as under implementation rather than completed. Future-oriented language is often hedged. This signals aspiration, but it is constrained by resource realities.

Figure 7 identifies the main sources of positivity in Malaysian Reddit’s Technological Innovation frame (overall sentiment: +0.41): Employment opportunities (+0.48), Corporate investment (+0.42), International partnerships (+0.38), and Technological advancement (+0.35). In contrast with Frame 1 with its individual negativity, Frame 2 is filled with optimism, which is spread among the institutional categories: users project Positive Affect to corporate action, cross-border cooperation, and proven technological benefits.

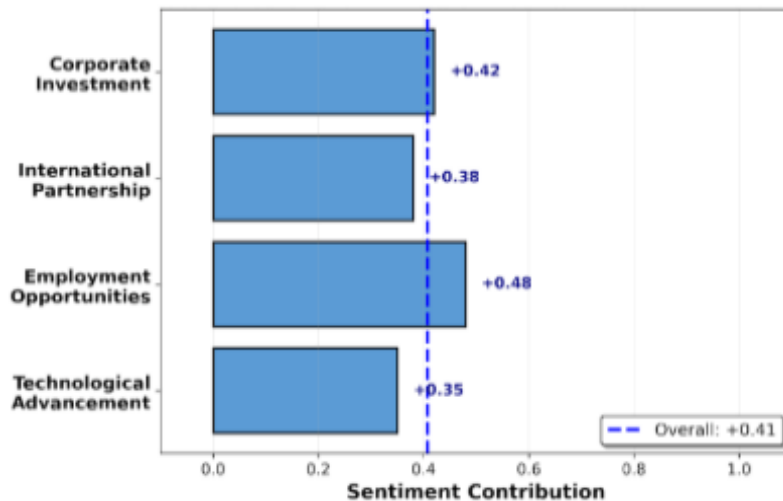


Figure 7. Linguistic positivity sources in frame 2 Reddit

Detailed sentiment statistics further show that Reddit’s acceptance of technology is relatively reserved and prudent. Positive sentiment accounts for 10.9%, while negative sentiment reaches 7.3%. The positive-to-negative comment ratio is 1.5:1. This pattern suggests that Reddit Malaysia tends to prioritize more contentious and critical discussions. Negative sentiment clusters around



concerns about national institutions and structural constraints. Typical themes include limited technological readiness in the Malaysian market, implementation challenges linked to local infrastructure, and brain drain affecting the retention of tech talent. Graduation in such corpus extends to scale markers technological scope, but also mentions dependency on external platform; participation is often made in conditionals that recognize practical limits; attitude is often made to appraise capacity, appreciative of technical ability and mobilizing resources.

In Frame 2, corporate-subject + transitive-verb + technological-object patterns frequently present leading technological firms as prime movers, with state agencies “coordinating” and local actors “adapting.” Comparative syntax (“catching up to...”) situates Malaysia relationally and can tilt the narrative toward latecomer progress. Taken together, these features portray AIED adoption as pragmatic and partnership-led, while also registering concerns about dependency and uneven rollout. When read alongside Frame 1 and Frame 3, this pattern suggests that coordination across funding, implementation, and employability merits closer, system-level attention.

Frame 3: Pedagogical Implementation- Market Navigation via Personal Credential Accumulation

Frame 3 draws on LDA Topic 4 and BERTopic Topics 0 and 2, comprising 31.1% (n = 184/598) of the Reddit corpus. Higher education discourse on Malaysian Reddit exhibits consistent grammatical instrumentalization of learning: education is linguistically positioned as a means to secure employment rather than as an end in itself. At the syntactic level, clauses link “training,” “degree”, and “course” with outcome verbs such as “get”, “find”, and “work”, producing cause-effect structures that naturalize education as labour-market preparation. Nominalization turns processes into measurable results such as “graduate employability,” “career readiness,” and “industry relevance.” It recasts pedagogy as performance indicators aligned with economic productivity.

Table 3. The distribution of top-30 most salient keywords for Frame 3

LDA	BERTopic
<p>Topic 4: teacher, university, course, education, model, student, teach, teaching, class, training, classroom, knowledge, learning, educational, study, history, time, new, middle, innovation, question, intelligent, year</p>	<p>Topic 2: campus, academic, training, qualification, engineering, faculty, higher, tertiary, institution, college, diploma, undergraduate, graduate, bachelor</p> <p>Topic 0 (partial): degree, university, computer_science, course, job, software, computer, science, student, uni, good, malaysia, study, program, major, education,</p>

As Figure 8 illustrates, higher education discourse effectively becomes employment discourse rendered in pedagogical vocabulary. Discussions of teaching, curriculum design, and assessment are framed by their contribution to job placement; industry-university collaboration serves as the principal metric of institutional success; and “career anxiety” functions as a proxy for educational quality. This language trend combines instruction with marketability, whereby academic



institutions and learners have a role to play in establishing a grammar of need where education is relevant to the market.

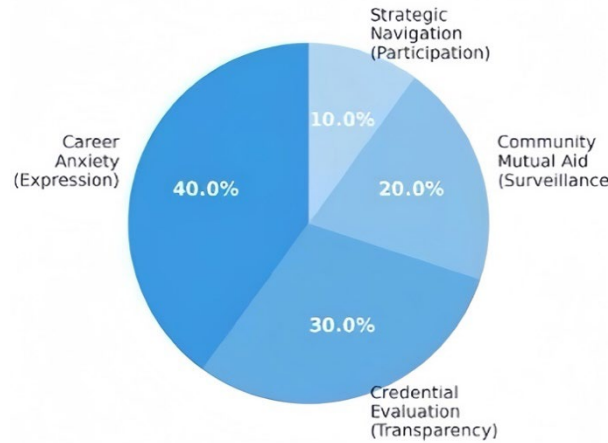


Figure 8. Functional composition in Frame 3

Pattern 1: [Interrogative clause: credential choice] + [community response: empirical information sharing] + [strategic calculation adjustment]

- Should I pursue master's or enter workforce? What starting salaries can local bachelor's expect?
- Does anyone else feel perpetually behind despite continuously upskilling?
- UM computer science vs. Monash Malaysian campus—which has better job placement rates and employer recognition?

As a grammatical form of educating the problem of risk management, which needs to be crowdsourced by peer intelligence, education is created through interrogative saturation (41.3% of Frame 3 discourse). This trend is indicative of the unique community design of Reddit: non-hierarchical communication that is anonymous and lacks social distancing considerations will allow functional inquiry functionality that will increase user engagement and result in the creation of a lively collaborative knowledge sharing experience. The facts that users desire to know include the salary rates, the cost of certifications, interview experience, and the reputation of companies that they cannot access or trust through the official resources, and therefore they exchange such with each other as a survival strategy.

This syntax generates a peer-support support framework in which the anxiety is directed to joint navigation instead of structured defiance. Community feedback allows users to improve strategic calculations by sharing the experience of the process into a distributed solution to all individuals who share their problem. Nevertheless, market logic is recreated in the form of the question. It both presents education as an individual path which is steered by peer advice, and not a place where coordination with peers or institutional responsibility occurs. The way Reddit is designed helps in mutual aid. It also normalizes the competition of credentials. It is a call to get used to things and not to reform the system.



Pattern 2: [Credential/Degree as carrier] + [intensive relational process] + [market value attribute] + [individual actor] + [material acquisition process] + [supplementary requirement]

- Computer science degree is insufficient—employers require React, cloud computing, and portfolio projects;
- Local university credentials are devalued—overseas qualifications command higher starting salaries;

Relational clauses (“is insufficient,” “has limited value,” “are devalued”) construe credentials as commodities with market attributes rather than outcomes of transformative learning. Degrees do not provide/are capabilities anymore, they are/have inspectable traits; education thus is subject to the test of the market. Employers occupy mental-process positions (“industry demands”), naturalizing the labour market as arbiter of value. Individuals appear as actors (“must acquire certifications/experience”), with listings depicting discrete investments and institutional mediation. Modality centres obligation on individuals, rendering credential inflation an individual responsibility.

Pattern 3: [Individual subject] + [modal obligation: must/need/should] + [material acquisition process] + [competitive objective]

- Students must acquire coding skills beyond curriculum to meet employer expectations for entry-level developer positions
- Job-seekers should continuously upskill through online courses and bootcamps to maintain employability in credential-inflated environment

The coupling of necessity modals (“must/need/should”) with individual subjects and acquisition verbs constructs credential supplementation as personal liability. Deontic force outweighs possibility, producing imperative urgency and an insufficiency cycle. Outcome clauses use probabilistic modality (“might secure interviews”), yielding a deontic–epistemic asymmetry: high effort, low-certainty returns. Responsibility shifts to individuals while institutional commitments recede. Accountability discourse, when present, names specific actors and quantifies deficits, centralizing responsibility through per-capita and regional comparisons.

Pattern 4: [Temporal marker] + [quantified credential formula] + [inequality operator] + [outcome negation]

- 10 years ago: bachelor’s degree = job offer; NOW: bachelor’s + master’s + 3 certifications + portfolio + 6-month internship ≠ guaranteed employment
- Previous generation: diploma sufficient for stable career; Current generation: bachelors minimum + continuous upskilling mandatory + perpetual inadequacy

While Frame 2 presents horizontal comparison across countries and regions, Frame 3 focuses on vertical comparison over time through mathematical formulations. The mathematical format (equations, inequalities, additive formulas) grammatically construes credential inflation as a calculable objective phenomenon, not a subjective perception or rhetorical complaint. They tabulate the numbers of credentials, comparisons, and time standards. This is a level of numeracy



which proves individual anxiety to be a rational reaction to a quantifiable system. Also, the inequality operator syntactically encodes perpetual insufficiency, in which the accumulation of credential by individuals can never be sufficient to produce affective orientation, that is one of perpetual inadequacy. This inadequacy is also strengthened by the temporal contrast structure where current is in comparison with the idealized past with its simpler qualifications that would yield improved results.

The pairing (Past: X sufficient; Now: X+Y+Z insufficient) is a syntactic juxtaposition constructing a story of devaluation in which the modern generation is burdened with more and more demands accompanied by less and less returns. The contrastive temporal marker fulfills both of its purposes: it confirms the present and constant state of anxiety by providing a historical reference point and renders inflation an irreversible process but not an error in policy.

The semantic rhythm created by the systematic co-occurrence of credential terms and insufficiency markers constructs an education discourse where mentioning a degree automatically triggers associations with inadequacy. This lexico-semantic valence generates the strong instructional negative sentiment (-0.736) because universities are implicitly blamed for issuing credentials deemed insufficient by the market; simultaneously, it produces the moderate career negative sentiment because inflation becomes a normalized, perpetual condition to be managed.

The emotional trajectory (see Figure 9) slides from strong instructional negative to moderate career negative, which is not an improvement in prospects but an accommodation to the “normalization of credential inflation”. The syntactic mechanism for this is a shift in responsibility and agent distribution. In the instructional discourse, the named actor + quantified deficit uses relational processes to tag the university as the accountable “underperformer”, creating a narrative of correctable fault and amplifying anger.

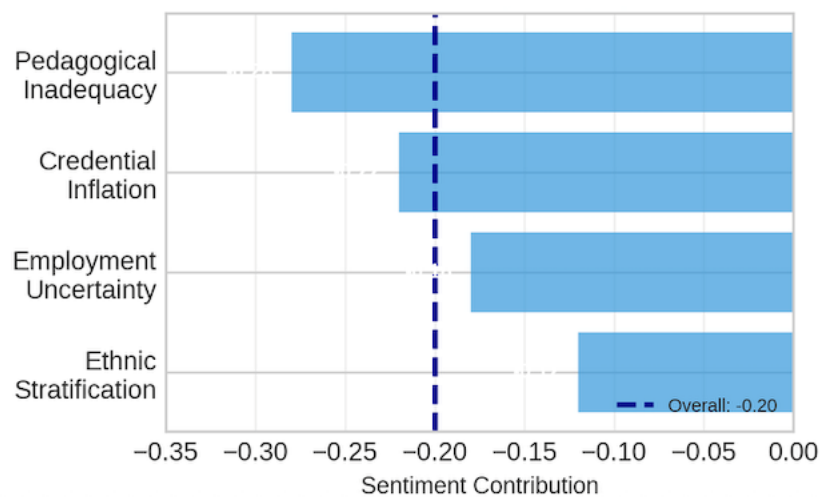


Figure 9. Negativity production mechanisms in frame 3

Temporal encoding couples a deontic present with probabilistic futures: simple-present assessments (‘skills gap widens’, ‘competition intensifies’) create immediacy, while modals and conditionals (‘will secure’, ‘might lead’) defer payoffs and make educational value contingent on labor-market validation. The result is urgency now and uncertainty later.



The affective pattern reveals strong negative sentiment toward instruction (-0.736) and moderate negative sentiment toward careers (-0.167), coupled with a complete absence of industry coordination discourse. Figure 9 reveals negativity production mechanisms (overall: -0.20): Pedagogical Inadequacy generates the strongest negative affect, followed by Credential Inflation, Employment Uncertainty, and Ethnic Stratification. Unlike Frame 1's personalized accountability or Frame 2's institutional optimism, Frame 3's negativity distributes across systemic failures.

The affective curve shifts off the high instructional negative to medium career negative indicating not improvement but fitting in to the normalization of credential inflation. The mechanism of this reversal consists in this redistribution of responsibility: instructional discourse positions the universities as irresponsible underperforming units, increasing the intensity of anger by producing narratives of correctable fault; career discourse spreads responsibility to the market, employers and macro conditions, breaking down definite objects of negative affect because of probabilistic modality.

The amplification of anxiety caused by graduating generates anxiety-inducing markers on an ever-competitiveness scale ("it has to keep upskilling", "continue to upgrade my skills", etc.), all of which makes the labour market look like a treadmill whose speed is ever-increasing. Engagement communicates dialogic space in terms of facts such as the employment, the law of market demands as something unnegotiable. Centre of attitude is focus on capacity anxiety. The basis of judgment on credentials is no longer on the quality of education but the ability of the credentials to result in the employment of the curriculum holder.

This constructs AIED meaning under a form of credentialist pragmatism: educational legitimacy is a manifestation of the demonstration of employability results, and negative affect is concentrated on pedagogical inadequacy and credential devaluation, which is an expression of a disorganized system in which the individual learners are tasked with negotiating market demands without organized institutional support to do so. Finally, the grammar of pedagogy closes the loop by placing the obligation to adapt on individuals, turning education into a perpetual condition of self-investment. This pattern is consistent with empirical findings showing that learners are expected to navigate AI tools independently, even when institutional support structures for ethical and effective use remain underdeveloped (Robert et al., 2025).

DISCUSSION

Synthesizing these findings, the three frames trace a single "through line": Malaysian Reddit constructs a coherent yet fragmented grammar of obligation in AIED. Frame 1 confines the state to calculable inputs ("allocates/provides/funds"); Frame 2 casts major technology firms as the movers of implementation; Frame 3 closes the loop by placing the obligation to adapt on individuals, turning education into a perpetual condition of self-investment. In this arrangement, no actor is linguistically accountable for linking Frame 1's investments, Frame 2's implementations, and Frame 3's employment outcomes. Coordination gaps are reinterpreted as students' navigational deficits rather than institutions' design failures.



Appraisal resources consolidate this settlement. Attitude centres on capacity anxiety, valuing credentials by placement yield rather than pedagogical quality. Engagement narrows debate through factual employment claims that render market demand non-negotiable. Graduation amplifies urgency through intensifiers such as “increasingly competitive” and “constant upskilling required,” which make continual self-improvement appear both rational and necessary. These co-occurrences are not accidental; they linguistically stabilize a worldview in which success is individualized and coordination is backgrounded.

A second, cross-frame finding is temporal disjuncture. Frame 1 speaks the iterative present and present perfect of budget cycles (“allocates,” “has increased”); Frame 2 favours the present continuous of projects (“implementing,” “testing”); Frame 3 pairs deontic presents with probabilistic futures (“must acquire now,” “might secure later”). The cumulative effect is permanent deferral: policies remain “upcoming,” pilots “in progress,” while students must decide within months under multi-year uncertainties. This misalignment normalizes precarity as a personal planning problem rather than a coordination failure. This temporal misalignment constitutes the central theoretical contribution of the study: it demonstrates that fragmented accountability is not only distributed across actors through transitivity patterns, but is also reproduced across time through tense and modality choices, making the grammar of obligation a structurally reinforced rather than incidental feature of AIED discourse.

The significance of this pattern lies in what it forecloses: when the grammar of all three frames consistently places risk on individuals, it becomes linguistically difficult to articulate shared institutional accountability, which has direct implications for how policy reform can be imagined and communicated.

Each frame is internally rational. Taken together, they normalize a settlement in which alternative grammars that assign shared responsibility for linking funding, technology, and employment become linguistically awkward and therefore politically unlikely. Recognizing this does not prescribe policy, but it clarifies why well-intentioned calls for “more funding,” “more partnerships,” and “more upskilling” tend to reproduce fragmented accountability.

CONCLUSION AND RECOMMENDATION

Naturally, the limitations of this study must be candidly addressed. Users of social media do not constitute an absolute reflection of the population of citizens as a whole; most of our data artificially amplifies the voices of the urban middle classes and technologically adept populations, leaving marginalized rural peoples, families with low incomes, and low technical literate behind or speaking obliquely. The decentralized nature of Reddit influences the characteristics of the data, possibly increasing the extremity of the opinion, and blocking others under the moderation processes of the community.

Still, more importantly, the cross-sectional data does not reflect temporal dynamics of the discourse development. More so, discourse analysis demonstrates macro-mechanisms of the construction of meaning, but cannot capture the micro-level workings of power. For instance, the study cannot address how individual teachers negotiate AI tools in their daily practice, how



students experience datafied monitoring psychologically, or how marginalized communities resist or reappropriate mainstream AIED discourse. This needs to be answered through ethnography. Discourse analysis should be augmented with field observation, in-depth interviews, and participatory action research. This needs to be answered through ethnography. Discourse analysis should be augmented with field observation, in-depth interviews, and participatory action research.

Nevertheless, despite these restrictions, the main messages of this research are clear: The findings indicate that assessments of AIED based solely on technical metrics, policy documents, or pilot project outcomes risk overlooking the discursive mechanisms through which accountability is distributed and precarity is normalized. Educational technology encodes particular values, power relations, and institutional assumptions, and these are visible in the syntactic and evaluative choices documented in this study. Recognising this offers a basis for more critically informed policy dialogue around equity, institutional responsibility, and the conditions under which AIED can serve broader educational goals.

REFERENCES

- Alfredo, R., Echeverria, V., Jin, Y., Yan, L., Swiecki, Z., Gašević, D., & Martinez-Maldonado, R. (2024). Human-centred learning analytics and AI in education: A systematic literature review. *Computers & Education: Artificial Intelligence*, 6, Article 100215. <https://doi.org/10.1016/j.caeai.2024.100215>
- AoIR. (2020). *Internet research: Ethical guidelines 3.0 association of internet researchers*. Association of Internet Researchers. <https://aoir.org/reports/ethics3.pdf>
- Bernama. (2024, January 16). *Govt serious about digital transformation, promoting AI literacy – PM Anwar*. Prime Minister’s Office of Malaysia Official Website. https://www.pmo.gov.my/en/news-en/govt-serious-about-digital-transformation-promoting-ai-literacy-pm-anwar/?utm_source
- Blei, D., Ng, A., & Jordan, M. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993–1022. <https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf>
- Brzustewicz, P., & Singh, A. (2021). Sustainable consumption in consumer behavior in the time of COVID-19: Topic modeling on Twitter data using LDA. *Energies*, 14(18), 1–20. <https://doi.org/10.3390/en14185787>
- Chiu, T. K. F., Ahmad, Z., Ismailov, M., & Sanusi, I. T. (2024). What are artificial intelligence literacy and competency? A comprehensive framework to support them. *Computers and Education Open*, 6, Article 100171. <https://doi.org/10.1016/j.caeo.2024.100171>
- Dai, X., & Pan, H. (2024). A comparative analysis of attitude resources in Chinese and American English news reports on “roe v. wade.” *Journal of Language Teaching and Research*, 15(2), 509–518. <https://doi.org/10.17507/jltr.1502.19>
- DeVito, P., Vallala, A., McMahon, S., Hinda, Y., Thaw, B., Zhuang, H., & Kalva, H. (2025). Unpacking Generative AI in education: Computational modeling of teacher and student perspectives in social media discourse. *IEEE Transactions on Computational Social Systems*, 1–13. <https://doi.org/10.1109/TCSS.2025.3630587>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *ArXiv*. <https://arxiv.org/abs/1810.04805>



- Fairclough, N. (2015). *Language and power (3rd ed.)*. Routledge.
- Fairclough, N. (2005). Discourse analysis in organization studies: The case for critical realism. *Organization Studies*, 26(6), 915–939. <https://doi.org/10.1177/0170840605054610>
- Grootendorst, M. P. (2024). *Semi-supervised topic modeling - BERTopic*. Maartengr.github.io. https://maartengr.github.io/BERTopic/getting_started/semisupervised/semisupervised.html
- Halliday, M. A. K. (1978). *Language as social semiotic: The social interpretation of language and meaning*. Edward Arnold.
- Ibrahim Brian, M. S. (2025). Reimagining English language learning: A systematic review of AI integration in classroom practice (2019–2024). *Journal of Creative Practices in Language Learning and Teaching*, 13(2), 14–29. <https://ir.uitm.edu.my/id/eprint/121086/>
- Jamaluddin, F., Jamaluddin, A. H., Jamaluddin, F., & Jamaluddin, F. (2025). *Malaysia & AI-driven education landscape: Policies, applications, and comparative insights for a digital future*. ArXiv.org. <https://arxiv.org/abs/2509.21858>
- Jessop, B. (2002). Liberalism, neoliberalism, and urban governance: A state-theoretical perspective. *Antipode*, 34(3), 452–472. <https://doi.org/10.1111/1467-8330.00250>
- Kamaruddin, S. S., Abdul-Rahman, S., & Wibowo, W. (2024). Understanding Malaysian public opinion on suicide through sentiment analysis and topic modeling of Reddit posts. *Engineering, Technology & Applied Science Research*, 14(6), 18055–18062. <https://doi.org/10.48084/etasr.8738>
- Koselleck, R. (2004). *Futures past: On the semantics of historical time* (K. Tribe, Trans.). Columbia University Press. (Original work published 1985)
- Lewis, A. A. (2025). Unpacking cultural bias in AI language learning tools: An analysis of impacts and strategies for inclusion in diverse educational settings. *International Journal of Research and Innovation in Social Science*, IX, 1878–1892. <https://doi.org/10.47772/IJRISS.2025.9010151>
- Lin, C.-J., Wang, W.-S., Lee, H.-Y., Li, P.-H., Huang, Y.-M., & Wu, T.-T. (2025). Advancing self-directed learning in STEM education: Integrating GPT-based learning aid with multimodal learning analytics. *Journal of Research on Technology in Education*, 1–19. <https://doi.org/10.1080/15391523.2025.2504358>
- Liu, Y., Duan, Y., Zhang, X., Dong, Y., & Hu, X. (2024). Have you been to Xiaohongshu today? Topic extraction and motivation analysis of user information sharing in online communities. *Documentation, Information & Knowledge*, 41(4), 110–120. <https://doi.org/10.13366/j.dik.2024.04.110>
- Mansoor, H., Bawazir, A., Alsabri, M. A., Alharbi, A., & Okela, A. H. (2024). Artificial intelligence literacy among university students — a comparative transnational survey. *Frontiers in Communication*, 9. <https://doi.org/10.3389/fcomm.2024.1478476>
- Martin, J., & White, P. R. R. (2005). *The language of evaluation*. Springer.
- Mimno, D., Wallach, H., Talley, E., Leenders, M., & McCallum, A. (2011, July 1). *Optimizing semantic coherence in topic models*. ACLWeb; Association for Computational Linguistics. <https://aclanthology.org/D11-1024/>
- Ministry of Education. (2023, November 28). *Digital education policy*. Ministry of Education. <https://anyflip.com/ncosr/fkhr/basic>
- Moe, T. M. (2015). Vested interests and political institutions. *Political Science Quarterly*, 130(2), 277–318. <https://doi.org/10.1002/polq.12321>



- Robert, J. D., Joseph, A. H., & Apolonius, L. E. (2025). AI-Assisted language learning in education: ESL perceptions and challenges of ChatGPT. *Journal of Creative Practices in Language Learning and Teaching*, 13(2), 48–65. <https://ir.uitm.edu.my/id/eprint/115953>
- Robinson, J. (2025). Beyond Koselleck's Neuzzeit: The deceleration of crises in late modernity Joe Robinson. *Journal of Political Ideologies*, 1–20. <https://doi.org/10.1080/13569317.2025.2518080>
- Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). *DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter*. Huggingface.co. https://huggingface.co/docs/transformers/en/model_doc/distilbert
- Schleicher, A. (2024). Toward the digital transformation in education. *Frontiers of Digital Education*, 1, 4–25. <https://doi.org/10.1007/s44366-024-0018-7>
- Stevens, D. (2021, February 19). *Fetching all comments from a Reddit post*. Gist. <https://gist.github.com/davestevens/4257bbfc82b1e59ecec7085e66314215>
- Suppiah, P. C., Noor, M. M., Turiman, S., Azizan, M., Aziz, M. N., & Bahrn, F. F. S. (2025). Behind the headlines: Modality in Malaysian news media on foreign labour recruitment. *Journal of Creative Practices in Language Learning and Teaching*, 13(2), 78–96. <https://ir.uitm.edu.my/id/eprint/121080>
- Taboada, M. (2016). Sentiment analysis: An overview from Linguistics. *Annual Review of Linguistics*, 2(1), 325–347. <https://doi.org/10.1146/annurev-linguistics-011415-040518>
- Verbytska, A. (2024). Topic modelling as a method for framing analysis of news coverage of the Russia-Ukraine war in 2022–2023. *Language & Communication*, 99(10), 174–193. <https://doi.org/10.1016/j.langcom.2024.10.004>
- Viberg, O., Cukurova, M., Feldman-Maggor, Y., Alexandron, G., Shirai, S., Kanemune, S., Wasson, B., Tømte, C., Spikol, D., Milrad, M., Coelho, R., & Kizilcec, R. F. (2024). What explains teachers' trust in AI in education across six countries? *International Journal of Artificial Intelligence in Education*, 35(3), 1288–1316. <https://doi.org/10.1007/s40593-024-00433-x>
- Ye, L. (2024). Wordless (Version 3.6.0). In *GitHub*. <https://github.com/BLKSerene/Wordless>.
- Zhang, E., Zhang, H., Yao, J., & Wang, S. (2024). The dynamic evolution and structural analysis of AIGC topics: A comparative study based on Weibo and Twitter. *Journal of Xi'an Jiaotong University (Social Science Edition)*, 44(3).

Declaration of Generative AI and AI-assisted Technologies in the Writing Process

During the preparation of this work, the authors used ChatGPT for polishing the language to enhance readability. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Conflict of Interest

The authors have no conflicts of interest to declare.



Acknowledgement

This research was supported by the University of Malaya, Faculty of Languages and Linguistics (UM FLL) Research Fund.

Authors' Contributions

Author 1(Xia Yutong): Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Investigation, Formal analysis, Data curation, Conceptualization.

Author 2 (Dr. Musaev Talaibek): Writing – review & editing, Supervision, Validation, Formal analysis, Data curation, Conceptualization.