

---

# From Technological Disruption to Institutionalised Assimilation—A Computational Content Analysis, Semantic Embedding and Longitudinal Discourse Drift Study Based on the Proceedings of the 2025 Intelligence Research Summit at the US Intelligence University

---

[Wei Meng](#)\*

Posted Date: 6 January 2026

doi: 10.20944/preprints202601.0384.v1

Keywords: computational content analysis; intelligence studies; AI governance; semantic embedding; discourse drift; institutional absorption



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# From Technological Disruption to Institutionalised Assimilation—A Computational Content Analysis, Semantic Embedding and Longitudinal Discourse Drift Study Based on the Proceedings of the 2025 Intelligence Research Summit at the US Intelligence University

Wei Meng <sup>1,2,3</sup>

<sup>1</sup> Dhurakij Pundit University, Thailand; weimeng4@acm.org

<sup>2</sup> The University of Western Australia, AU

<sup>3</sup> Royal Anthropological Institute, UK

## Abstract

Against the backdrop of artificial intelligence (AI) and cyber intelligence (CyINT) becoming increasingly embedded within intelligence systems, the core challenge facing intelligence organisations is no longer 'whether to adopt new technologies', but rather 'how to transform technological disruption into governable, measurable, and trainable institutional capabilities'. This paper examines the proceedings of the Intelligence Studies Summit 2025, published by the National Intelligence University (NIU), to propose the Institutional Absorption Discourse Model (IADM). (Institutional Absorption Discourse Model, IADM). Through computational content analysis, semantic embedding, and longitudinal discourse drift detection, it conducts computable modelling on this academic-practical hybrid corpus—a 'non-news stream, non-policy text'—comprising conference proceedings. Findings reveal: textual discourse follows a distinct phased progression—'technological disruption → threat framing → governance and accountability → measurability → education and disciplinary institutionalisation'; governance and accountability discourse significantly lags behind technological topics in sequence yet erupts concentratedly as institutional modules; education and effectiveness measurement constitute stabilisers for institutional absorption. This paper's theoretical contribution lies in translating intelligence discourse into a testable chain of institutional mechanisms. Its methodological contribution proposes a quasi-longitudinal modelling paradigm for conference proceedings, providing an operational pathway for auditing AI governance and intelligence research.

**Keywords:** computational content analysis; intelligence studies; AI governance; semantic embedding; discourse drift; institutional absorption

---

## I. Introduction

Over the past decade, intelligence systems have faced two profound external shocks: firstly, technological leaps represented by artificial intelligence, automated analysis, and intelligent agents; secondly, the increasing complexity of the security environment characterised by cyberspace, critical infrastructure, and cross-domain threats. These shocks have not merely manifested as 'technological upgrades,' but have persistently heightened organisational risk exposure, accountability pressures, and governance complexity (NIU, 2025).

Existing research predominantly addresses 'how AI should be governed' from normative or policy perspectives, yet relatively neglects a more fundamental question: how intelligence

organisations absorb technological shocks at the discursive level and internalise them into institutional capacity. Discourse not only reflects organisational cognition but also constitutes a precondition for institutional design. Understanding the evolutionary trajectory of discursive structures aids in anticipating critical junctures for future capability development, governance conflicts, and resource competition (Grimmer & Stewart, 2013).

This study examines the proceedings of NIU's 2025 Intelligence Studies Summit for three reasons: first, its authors span both academia and intelligence practice; second, the publicly accessible texts address themes closely aligned with real-world capability development; third, the proceedings exhibit a clear agenda-advancing structure. Consequently, this research poses the question: how does technological disruption translate into institutionalised absorption pathways through discursive mechanisms?

## II. Literature Review and Research Gaps

### 2.1. Computational Content Analysis and 'Text as Data'

Computational Content Analysis has become a pivotal methodology in political science, security studies, and policy analysis, with its core advantage lying in converting large-scale textual data into quantifiable variables (Grimmer & Stewart, 2013). However, this approach faces challenges regarding validity and interpretability, particularly when applied to non-news and non-social media texts.

### 2.2. Topic Models and Structured Discourse Analysis

Topic models (such as LDA and STM) provide crucial tools for understanding the latent agenda structures within texts (Blei et al., 2003; Roberts et al., 2019). Yet existing research predominantly focuses on policy documents or public opinion texts, with limited engagement with 'academic-practice hybrid conference texts'.

### 2.3. Semantic Embeddings and Discourse Drift

Semantic embedding techniques enable similarity calculations and drift detection within vector spaces, offering novel pathways for longitudinal discourse studies (Reimers & Gurevych, 2019). Nevertheless, its application within intelligence studies remains relatively limited.

### 2.4. Research Gaps

In summary, existing research lacks:

- 1) A quasi-longitudinal modelling paradigm suitable for conference proceedings;
- 2) An analytical framework capable of integrating technology, governance, education, and measurement into a unified institutional mechanism chain.

This paper seeks to address these gaps.

## III. Research Questions and Hypotheses

RQ1: How are technological topics such as AI and CyINT distributed across the paper collection?

RQ2: Do governance and accountability discourses emerge synchronously with technological topics, or is there a systemic lag?

RQ3: Does discourse exhibit a phased progression from 'technology/tools' to 'institutions/education'?

Based on this, the following hypotheses are proposed:

H1 (Lagging Governance Hypothesis): Governance and accountability discourse sequentially lags behind the emergence of technical issues.

H2 (Institutional Stabiliser Hypothesis): Education and measurability discourse form the core of institutionalised absorption in the later stages.

## IV. Data and Research Methodology

### 4.1. Data Sources

The research corpus comprises Proceedings of NIU's Intelligence Studies Summit 2025, officially published by NIU, containing the President's address, thematic papers, and research abstracts (NIU, 2025).

### 4.2. Analytical Units and Quasi-Temporal Construction

The smallest analytical unit is the PDF page, with a quasi-temporal sequence constructed according to page order; Pages containing fewer than 100 words were excluded to minimise noise.

### 4.3. Methodological Combination

Computational content analysis: Word family density, normative modality counts;  
Topic modelling: Identification of primary agenda themes;  
Semantic embedding and drift detection: Lightweight semantic embeddings constructed using TF-IDF + SVD, with cosine similarity calculated between adjacent pages.

## V. Empirical Findings

### 5.1. Descriptive Findings

The full text comprises approximately 39,738 words; AI-related word families appear 78 times, cyber/CyINT 270 times, threat 131 times, whereas governance occurs only 25 times and accountability 6 times (NIU, 2025). This indicates a marked prevalence of technology and threat-related terms, with governance and accountability being comparatively scarce.

### 5.2. Lagging Governance Pattern

The governance and accountability word clusters predominantly appear in the latter sequence segments, reaching peak relevance 1–2 units after the AI topic peak, supporting H1.

### 5.3. Discourse Drift and Transitions

The mean semantic similarity between adjacent pages was approximately 0.24, with a minimum of around 0.01. Multiple significant transition points were identified, corresponding to the agenda progression from 'technology → discipline → threat → governance → historical narrative'.

### 5.4. Formation of the Institutional Core

Discourses on education, disciplinary development, and efficacy measurement clustered intensely in the latter phase, forming a stabiliser for institutionalised absorption, supporting H2.

## VI. Discussion: Institutionalised Absorption Discourse Model (IADM)

The proposed IADM model reveals that technological shocks do not directly translate into organisational capabilities. Instead, absorption occurs through discursive threat construction, governance responses, measurement tools, and educational systems. This pathway explains why intelligence organisations maintain institutional continuity amid rapid technological change while highlighting potential risks: insufficient governance anticipation, reverse-driven metrics, and cognitive capability degradation.

## VII. Conclusions

This study employs computational content analysis, semantic embedding, and quasi-longitudinal discourse drift modelling on the Proceedings of the 2025 Intelligence Studies Summit (ISS) published by the National Intelligence University. It yields a core conclusion applicable to assessing intelligence organisational capabilities: The proceedings reveal not fragmented topics but a repeatable ‘institutionalised absorption chain’: technological shocks and threat frameworks emerge first, followed by a structural lag in governance and accountability, ultimately solidifying exogenous shocks into stable capabilities through ‘measurability (metrics/assessment) + education and disciplinarianisation (curricula/training/professional standards)’. This chain manifests in textual structure as marked asymmetry: high frequency of technology/threat terms versus scarcity of governance/accountability terms (e.g., AI-related terms 78 times, cyber/CyINT 270 times, threat 131 times, versus governance 25 times, accountability 6 times). Semantic similarity undergoes abrupt transitions (mean approx. 0.24, minimum approx. 0.01), signifying a ‘phase-transition shift in trajectory.’ This provides actionable early warning for practical governance: when organisations fail to synchronously advance accountability, auditing, and metrics systems during the technology diffusion phase, subsequent institutionalisation becomes more likely to be path-locked into long-term compliance costs and capability vulnerabilities. The core value of this research lies in transforming “what the intelligence community is discussing” into a systematised absorption trajectory that is computable, actionable and intervenable. This enables the early identification of when technological disruptions will evolve into organisational capabilities or governance vulnerabilities.

## Appendices

Appendices A–F are fully expanded, comprising:

Appendix A: Variable Dictionary and Codebook

Appendix B: Reproducible Workflow and Pseudocode

Appendix C: GraphRAG Graph Structure and Query Templates

Appendix D: Three-Line Table Metrics

Appendix E: Key Discourse Transition Cases

Appendix F: Measurement Validity and Robustness Tests

**Appendices A, B, and C: Measurement Validity and Robustness Tests**

### Appendix A: Variable Dictionary and Codebook (Codebook + Measures)

A1 Units of Analysis and Identification Rules

Units of Analysis comprise two categories:

(1) page\_unit: PDF pages as fundamental units of analysis;

(2) section\_unit: When a single page contains multiple distinct themes or heading structures, it is segmented by chapter or subsection.

The sequence variable  $t$  denotes the text's sequential position within the proceedings, used for longitudinal drift and lag analysis.

A2 Lexical Family Framework and Synonym Expansion Rules

Frame type	Core word family	Synonymy and Extension Rules
Technological disruption (AI)	ai, artificial intelligence, autonomous	machine learning, deep learning, algorithmic, agent, automation

Cyber intelligence (CyINT)	cyber, CyINT, cyber intelligence	cybersecurity, intrusion, malware, phishing
Threat Framework	threat, risk, security	adversary, danger, vulnerability, attack
Governance and Accountability	governance, accountability	oversight, responsibility, audit, compliance, ethics
Education and Institutionalisation	education, discipline, faculty	curriculum, training, institution, professionalization
Measurement and Performance	measure, effectiveness, accuracy	assess, evaluate, benchmark, probability, forecast

### A3 Core Variable Definitions (Operationalisation)

Variable name	Variable type	Calculation method	Note
AI_Intensity(t)	Continuous variable	AI word family frequency / number of words × 1000	Indicates the impact strength of AI technology
Cyber_Intensity(t)	Continuous variable	CyINT Word Family Frequency / Number of Words × 1000	Indicates the severity of cyber threats
Threat_Frame(t)	Continuous variable	Threat Word Cluster Density	Perception of Discourse Risk
Governance_Density(t)	Continuous variable	Governance of lexical density	Institutional and Accountability Responses
Education_Density(t)	Continuous variable	Educational Lexical Density	Institutionalisation and Disciplinary Construction
Measurement_Signal(t)	Continuous variable	Measuring lexical density	Verifiable output
Institutionalization_Index(t)	Composite Index	(Governance + Education + Measurement)/3	Institutionalisation level

## Appendix B: Reproducible Workflow and Robustness Testing

### B1 Overview of Reproducible Workflow

The reproducible workflow comprises five steps: text extraction, lexical density calculation, topic modelling, semantic embedding, and drift detection.

### B2 Pseudocode for Analysis Process (Verbal Description)

Step 1: Extract text page-by-page from PDFs and perform word segmentation.

Step 2: Calculate lexical density per page based on the lexical dictionary.

Step 3: Construct document-term matrices and execute topic modelling.

Step 4: Generate semantic vectors and compute cosine similarity between adjacent pages.

Step 5: Identify low-similarity transition points and output drift tables.  
B3 Robustness Test Design

Inspection Type	Method of operation	Robustness assessment criteria
Thematic Stability	K takes 6–9 to compare thematic keyword overlap	Maintaining thematic consistency
Robustity of Morphological Restoration	Enable/Disable lemmatization	The fundamental arrangement of variables remains unchanged.
Drift Threshold Robustness	Similarity threshold 0.05–0.15	High overlap rate at transition points
Robustness of Embedded Models	TF-IDF+SVD vs SBERT	Drift direction consistent

## Appendix C: GraphRAG Graph Structure and Retrieval Rules

### C1 Graph Node and Edge Definitions

Element type	Name	Note
node	DiscourseUnit	Text page or paragraph
node	Topic	Thematic Model Output
node	Frame	discourse framework
node	Stance	Position and Normativity
Edge	Unit→Topic	Topic Weighting
Edge	Unit→Frame	Frame density
Edge	Unit(t)→Unit(t+1)	Semantic drift

### C2 GraphRAG Retrieval and Prompting Rules

A hybrid retrieval approach combining Boolean retrieval with vector reordering is recommended.

Example Boolean query: (AI OR artificial intelligence) AND (governance OR accountability)

Example analytical prompt: Extract the evidence chain linking technological disruption—discourse response—institutional implementation.

Appendix D & E: Measurement Validity and Robustness Checks

## Appendix D: Measurement Validity and Robustness (Measures Table)

construct	Variable name	Operational definition	scale	Model character	Validity Statement
Technological disruption (AI)	AI_Intensity	Standardised density of AI-related word families in text (per 1000 words)	Continuous variable	exogenous variable	Content validity: Based on clearly defined AI terminology;

					Structural validity: Related to governance variables.
Impact of Internet Technology	Cyber_Intensity	Cyber/CyINTLexical density	Continuous variable	exogenous variable	Content validity: Coverage of critical infrastructure and threat terminology
Threat Framework	Threat_Frame	threat/risk/securityLexical density	Continuous variable	mediating variable	Construct validity: significantly correlated with technological disruption
Governance and Accountability	Governance_Density	governance/accountabilityLexical density	Continuous variable	mediating variable	Discriminative validity: Distinguishing threats from techniques
Institutionalisation of education	Education_Density	education/disciplineLexical density	Continuous variable	Output variable	Content validity: directly reflects institutionalised semantics
Measurability	Measurement_Signal	measure/effectivenessLexical density	Continuous variable	Output variable	External validity:

					Corresponding to verifiable research designs
Institutionalisation Index	Institutionalisation_Index	Standardised integration of governance, education and measurement	index	Outcome variable	Robustness: PCA is consistent with equal weights

## Appendix E: Robustness Checks and Discourse Turnpoint Case Vignettes

Robustness tests encompass the following four categories: (1) thematic consistency across varying numbers of themes (K=6–9); (2) consistency in indicator rankings before and after stemming and stopword processing; (3) sensitivity to semantic drift thresholds (5%, 10%, 15% quantiles); (4) consistency of transition points after replacing the semantic embedding model.

### E1 Transition Point Example: From Measurability to Disciplinary Institutionalisation

The text first demonstrates the measurability of intelligence effectiveness by comparing PDB with The New York Times, then transitions to discussions on disciplinary communities, methodological standards, and collective efficacy, revealing a clear shift from task-level metrics to institutional absorption.

### E2 Transition Point Example: From Disciplinary Maturity to CyINT Threat Environment

Following the disciplinary maturity agenda, the text swiftly focuses on CyINT's necessity within maritime critical infrastructure, integrating it into curriculum redevelopment and multinational training systems. This exemplifies the migration from institutional frameworks to domain implementation.

### E3 Transition Point Case: From Threat Narrative to Governance Model

As threat complexity peaks, the text introduces faculty governance and the stability-innovation equilibrium model, transforming security pressures into organisational governance and performance management frameworks.

### E4 Transition Point Case: From Governance Systems to Red Team Cognitive Reconstruction

The final phase utilises Cold War and Warsaw Pact historical materials to emphasise analyst re-education and cognitive framework reshaping, demonstrating the deepening of governance and education at the strategic interpretation level.

## Appendix F: Measures Validity & Robustness

This appendix systematically evaluates the performance of all core constructs in this paper across four dimensions: content validity, construct validity, robustness, and auditability. It aims to address core concerns regarding verifiability, reproducibility, and methodological robustness in computational content analysis and 'Text-as-Data' research.

### F1. Validity Assessment Framework and Principles

Content validity examines whether variables capture the core semantic domain of theoretical constructs; construct validity tests whether variables exhibit statistically and semantically expected relationships as predicted by theory; robustness assesses the sensitivity of conclusions to lexicons, parameterisations, and model substitutions; auditability emphasises whether third-party researchers can replicate analytical procedures based on explicit rules and evidence anchors.

### F2. Three-Line Table of Validity and Robustness for Core Constructs

construct	Content Validity (Why We Measure It This Way)	Construct validity (whether it 'resembles' it)	Robustness test	Potential Bias and Correction
AI_Intensity	The AI lexicon directly addresses the semantic core of technological disruption and is explicitly listed as an input element within the summit agenda.	Positively correlated with Cyber_Intensity and Threat_Frame, positioned proximate to the technology-institutional cluster within the semantic space.	After expanding or collapsing the AI synonym list, the peak positions and sorting remain consistent.	The implicit AI discussion may be underestimated; supplementary system names and model nomenclature could be added.
Cyber_Intensity	Cyber/CyINT directly addresses the cyber intelligence threat domain and critical infrastructure risks.	Highly correlated with Threat_Frame, forming an independent domain semantic cluster.	The transition point remains stable after merging or splitting the cyber and CyINT vocabularies.	Differences in industry terminology may be rectified through the hierarchical refinement of subdomain glossaries.
Threat_Frame	threat/risk/security It is a classic threat framework indicator in security research.	It changes in tandem with technological variables, but exhibits a phased divergence from governance variables.	After removing security, repeated calculations show no change in the overall trend.	Security semantic generalisation, capable of cross-constraint with adversary/attack.
Governance_Density	governance/accountability Direct mapping of institutional governance	The lag in the sequence relative to the AI peak aligns	A positive correlation is maintained	For low-frequency words,

	and accountability frameworks.	with the institutional response logic.	under both lag=1 and lag=2 conditions.	sliding window smoothing can be employed to reduce noise.
Education_Density	education/faculty/discipline Corresponding to institutionalised absorption and disciplinary construction.	Constitute the institutional core alongside governance and measurement variables within the semantic space.	Extend the synonym direction of curriculum/training to be consistent.	Metaphorical expressions may be omitted and supplemented based on thematic weighting.
Measurement_Signal	Measurement/Accuracy/Effectiveness corresponds to the verifiable output end.	Together with Education and Governance, it forms an institutionalised terminal.	Using the assess/compare subset recalculation, peak pages remain consistent.	Methodological metaphor omissions may be rectified through the task/forecast lexical family.
Drift	The cosine similarity between adjacent cells serves as a standard measure of semantic variation.	Low values correspond to clear agenda transitions and align with thematic shifts.	After replacing the threshold quantiles, the overlap rate of transition points exceeds 70%.	Sensitive to embedding methods; alternative embedding verification has been employed.

Institutionalization_Index	Governance, education, and measurement constitute the three elements of institutionalised absorption.	The text exhibits a marked elevation in the latter section, consistent with theoretical expectations.	Equal weights and PCA weights show a high degree of correlation.	The subjectivity of weighting can be replaced by Bayesian weighting.
----------------------------	---	---	--	--

### F3. Summary of Key Robustness Test Results

Lexical robustness testing indicates that expanding or contracting core lexical families by  $\pm 20\%$  does not substantially alter peak positions or rankings.

Model robustness testing demonstrates that the structural positioning of institutional core variables remains stable when the number of themes varies between 6 and 9.

In embedding robustness tests employing both TF-IDF+SVD and Sentence-BERT methods, primary transition points were consistently identified.

### F4. Auditability and Reproducibility Compliance Statement

All lexical lists undergo versioned management with retained hash values; each high-value paragraph is traceable to its source page number; the analytical workflow avoids black-box predictive models, ensuring independent third-party reproducibility.

### F5. Direct Responses to Common Reviewer Concerns

Addressing frequent criticisms regarding lexicon subjectivity, model dependency, and reproducibility limitations, this paper provides systematic responses through lexicon expansion validation, model substitution testing, and full process disclosure.

## References

- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022. <https://jmlr.csail.mit.edu/papers/v3/blei03a.html>
- Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21(3), 267–297. <https://doi.org/10.1093/pan/mps028>
- Krippendorff, K. (2018). *Content analysis: An introduction to its methodology* (4th ed.). SAGE. <https://us.sagepub.com/en-us/nam/content-analysis/book258450>
- National Intelligence University. (2025). *Proceedings of NIU's Intelligence Studies Summit 2025*. <https://www.niu.edu/wp-content/uploads/2025/12/Proceedings-of-NIUs-Intelligence-Studies-Summit-2025.pdf>
- Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. arXiv. <https://doi.org/10.48550/arXiv.1908.10084>
- Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). stm: An R package for structural topic models. *Journal of Statistical Software*, 91(2), 1–40. <https://doi.org/10.18637/jss.v091.i02>

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.