Methodology

This impact assessment applied a natural language processing (NLP) pipeline to compare the *Global Solutions Summit (GSS) 2025 Action Points* with the *T20 Communiqué Recommendations*. The approach first established sentence-level semantic alignments between the two documents and then applied a structured set of policy themes to assign labels, enabling interpretability of the matches.

Text Preprocessing

Both documents were ingested in ".docx" format and segmented into units suitable for cross-comparison. The GSS document was parsed into clusters and action points, reflecting its organizational structure, while the T20 communiqué was tokenized into sentences, with section headers retained for context. Standard normalization (e.g., whitespace harmonization, duplicate removal) was applied.

Semantic Representation with Sentence-BERT

Each action point and T20 sentence was represented as a dense vector embedding using Sentence-BERT (SBERT) (Reimers & Gurevych, 2019). SBERT is a transformer-based language model derived from BERT (Devlin et al., 2019) but optimized for semantic similarity tasks through a *siamese* training architecture. Whereas BERT embeddings are context-sensitive but not directly comparable at the sentence level, SBERT produces sentence vectors positioned in a high-dimensional space such that semantically similar sentences lie close together.

In this study, we used the *all-MiniLM-L6-v2* variant. To compare embeddings, we applied cosine similarity, a metric that measures the cosine of the angle between two vectors. Cosine similarity ranges from –1 (opposite meaning) to 1 (highly similar), and is widely used in NLP as it captures semantic closeness independently of vector magnitude.

Cross-Document Alignment

Using the SBERT embeddings, we computed pairwise cosine similarities between all GSS action points and T20 sentences. For each action point, the three most similar T20 sentences were retrieved, forming candidate alignments across the two documents.

Thematic Framework and Topic Labeling

To make the cross-document alignments interpretable, we developed a thematic framework consisting of major policy domains such as *Equity*, *Digital Inclusion*, and *Climate Finance*. Each theme was represented by a curated set of key terms that are commonly used to signal that policy area.

To represent each theme in the embedding space, we computed a mean embedding: the average of the embeddings of all its key terms. This mean embedding functions as a central reference

point that captures the overall meaning of the theme. For instance, the theme *Equity* might be defined by terms such as *equity*, *fairness*, and *equality*; the average of their embeddings produces a vector that reflects the collective meaning of the theme.

Text units (action points and T20 sentences) were then assigned to themes using a hybrid method:

- Semantic assignment: Each sentence embedding was compared with the theme mean embeddings using cosine similarity. Themes with the highest similarity scores (above a fixed threshold) were assigned, allowing sentences to be linked to themes even if they did not use the exact keywords.
- Lexical assignment: In parallel, regular expressions were used to detect explicit occurrences of the key terms, ensuring that direct mentions were captured with precision.

For each action point–recommendation pair retrieved in the alignment stage, their assigned themes were compared and overlapping domains were identified. This process enriched the similarity matches with substantive policy context, highlighting shared policy themes between GSS and T20 recommendations.

Output

The results were consolidated into a structured dataset containing GSS action points, their top-3 T20 matches, assigned topics, and overlapping themes.

Bibliography

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 3982–3992. Association for Computational Linguistics. https://doi.org/10.18653/v1/D19-1410

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, 4171–4186. Association for Computational Linguistics. https://doi.org/10.48550/arXiv.1810.04805