

Staying in the Loop with Gen AI: AI/Gen AI-Powered HLT for the Public Sector

Konstantine Boukhvalov

ManpowerGroup Public Sector, Inc.



ManpowerGroup™

ManpowerGroup Public Sector

- **30+** years supporting commercial and government clients in over 200 languages and dialects
 - Translation and I10n, transcription, interpretation, language technology support, linguist placements, multimedia analysis and reporting
- **Super user of Human Language Technology (HLT)**
 - 25+ years customizing language automation/HLT tools in 60 languages to achieve efficiencies, process voluminous materials, and provide cost savings
 - Leverage and adapt commercial products and combine tools to optimize technology to best meet customers' needs:
 - Open-Source Intelligence (OSINT), Translation Management Systems (TMS), Machine Translation (MT), CAT/Localization Tools, Authoring, eLearning, Desktop Publishing/Graphics Design, Audio/Video Production, Lexical Data Management, Optical Character Recognition

Objective

- **Vision Outline:**

- Evolving role of “humans-in-the-loop” for MT and Translation-related Services (TRS) for Public Sector support
- Organizational transformation related to Gen AI technologies and workflows

- **Focus Areas:**

- Corpus building
- Corpus curation / Quality control
- Security
- Workflow adjustments
- Output quality evaluation, including fact-checking and domain-specific expertise



Key Takeaways

- Defining the ever-evolving human role in MT and Translation-related Services (TRS) for Public Sector support
- Organizational and team transformation for “Gen AI success”
- Methods for building, optimizing, and securing ML corpora for MT and Gen AI training
- How to incorporate Gen AI into the workflows
- SME role in Gen AI output quality assessment



Gen AI – Translation/TRS Use Cases

- Translation
- Post-editing
- Collection / source identification
- Transcreation
- Simship – simultaneous content creation in multiple languages
- Summarization/roundups
- Transcription
- Multimedia preprocessing and intelligence (indexing, STT, entity recognition)
- Digitization
- Sentiment and emotion analysis



Gen AI vs. Humans: Better Translation and Faster Analysis?

- **An intelligent entity or a tool?**
 - Only as good as the training corpus
 - Relies on data collection
 - Requires curation
 - Bias and hallucinations
 - Requires privacy and security enforcement:
 - Training corpus
 - Prompts
 - Aux technologies (e.g., retrieval-augmented generation (RAG), agents)



Data Discovery and Collection

Gen AI/ML TR/TRS training corpora sources and human expertise requirements:

- Existing Translation Memories – Linguist + SME
- Non-aligned legacy translations with accompanying source – SME
- Monolingual content – SME
- Bilingual/multilingual open-source content – Linguist + SME
- Multimedia collection segments – Linguist + SME



Training Corpus Generation

Domain-specific AI/ML corpus generation for translation and TRS methods and techniques:

- MT
- PEMT
- Alignment
- Synthetic translation
- Bilingual conversion
- Corpus optimization and quality control



Outcome and Invested Effort:

- Results – from gold-standard to acceptable/required quality
- Human technologist and SME – involvement varies greatly depending on the selected approach

Data Privacy and Security

Data sharing – a key component and a key concern:

- Security analysts need to assess the data privacy and security supported by specific Gen AI providers – including custom LLMs – and continuously monitor for any changes in SLAs.

Managing sensitive data (CUI, PII, PHI, etc.) in AI/Gen AI output:

- Data removal/isolation in training corpora
- Filtering data via prompts
- Utilizing an auxiliary knowledge base / retrieval-augmented generation (RAG)



Workflow Augmentation

Challenges (examples):

- PEMT/MTPE – with the introduction of Gen AI, expected efficiencies
- Transcription – sensitive data transcription required low level of technology and high level of human involvement
- Audio/Video Intelligence – deep human involvement, including translation and analytics



Workflow Augmentation (cont.)

Solutions:

- PEMT/MTPE – post-NMT Gen AI review followed by human fact-checking
 - Time savings are still in evaluation
- Transcription – WAVpedal human transcription replaced by AI/Gen AI
 - Translator, transcriber, quality control roles merged into a single role – Linguist Fact-checker
 - New role required technology training
 - Time savings of over 60% vs. traditional workflow



Workflow Augmentation (cont.)

Solutions (cont.):

- Audio/Video intelligence – human video analysis replaced by AI/Gen AI supporting automated transcription, timing, grouping, speaker identification and separation, translation, key frame/object/entity/activity identification, and source/target-language subtitling
 - Translator, transcriber, analyst, and quality control roles merged into a single role – Linguist-analyst
 - New role required technology training
 - Time savings of over 80% vs. traditional workflow



Data Management Discipline

Data management discipline required to more effectively support current AI initiatives and transition to GAI-powered workflows:

- Identify and separate training corpora based on defined parameters (developed internally, provided by client, source, domain, etc.)
- Verify and unify corpora formats and schemas (as applicable)
- Create a multivector datastore with indexed, metadata-enriched resources (e.g., linked video, extracted key frames, transcriptions, translations, summaries, etc.)
- Identify and isolate sensitive data
- Continuously monitor data development volume and quality



Don't Just Stay in the Loop – Manage the Loop

- **Human expertise** – key to implementing successful Gen AI-supported production workflows.
- What should the “workforce of the future” look like?
 - Trained technologists who organically added prompt engineering to their existing advanced computer skills

OR

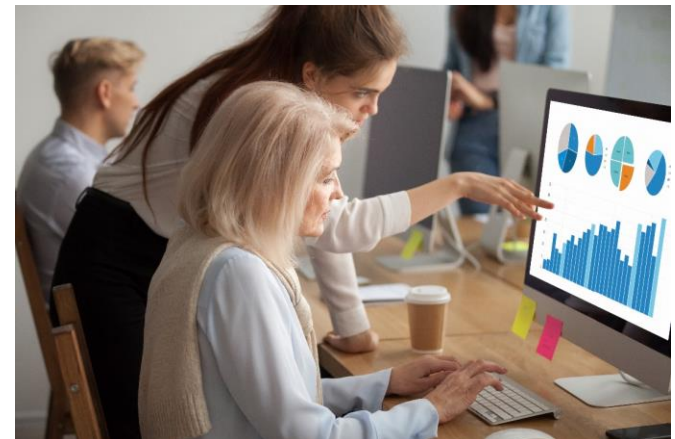
- Subject-matter experts who require more extensive training in technology but possess vast domain-specific knowledge



Don't Just Stay in the Loop – Manage the Loop (cont.)

The answer is both!

- Developing the initial ability to generate prompts takes weeks
- Honing those skills takes months
- Developing the subject-matter expertise required to construct a focused query takes years, possibly decades
- Strong support from engineers is required to:
 - Complement SMEs' newly acquired Gen AI expertise
 - Provide in-depth technology support and
 - Stay abreast of the new developments



Looking into the Future

- Vast expansion of Gen AI TRS support
- Multidisciplinary integration
- Automated Gen AI-powered workflows





ManpowerGroup

Thank you

Konstantine G. Boukhvalov

Director, Human Language Technology

ManpowerGroup Public Sector

+1-703-245-9372

Konstantine.Boukhvalov@manpowergroupsecure.com