AIDA Agents: A Multi-Agent Translation Platform with Context-Aware Quality Control

DATAmundi.ai

Emanuele Di Rosa · Piotr Peszynski · DATAmundi.ai

EAMT 2026

Tilburg

The problem

Enterprise localization faces a trade-off:

- **NMT engines** — need expensive *fine-tuning* and heavy human post-editing
- **LLMs** — high quality, but no localization workflow / CAT integration
- **Research multi-agent systems** — promising, but not production-integrated

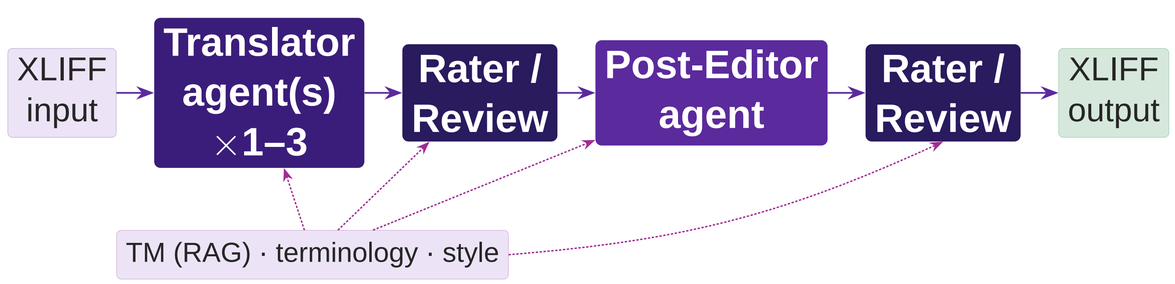
Need: high quality *and* drop-in deployment, with no fine-tuning.

Contribution: AIDA Agents

A *deployed* multi-agent LLM translation platform that:

- Orchestrates specialized agents — translation, rating, post-editing, re-rating
- Needs **no prior fine-tuning**; adapts via optional **RAG** (translation memories, terminology, style guides) injected at *every* stage
- Integrates **natively with XLIFF / CAT** tools; works with any NMT or LLM backend
- Is **deployed in production** at DATAmundi for enterprise clients

System architecture



Translator agents produce candidates; an MQM-inspired **Arbiter/Rater** scores each (accuracy, fluency, terminology, style) and selects the best with an explainable report; a **Post-Editor** refines it; a second rater produces the final XLIFF. Document context is present at *every* stage with ”batch segment processing”.

Three product versions

Version	Calls	What
AIDA LLM	3	1 translator + 1 post-editor + 1 rater — best quality/cost
AIDA MultiLLM	≤6	up to 3 parallel translators + rater + post-editor + re-rater — maximum quality
AIDA NMT	3	1 NMT + LLM post-editor + LLM rater — lowest cost

Why it works

- **Diversity + selection** — multiple translator agents, an MQM-style arbiter picks the best candidate
- **Explainable QE** — a weighted accuracy / fluency / terminology / style report drives a conservative post-edit
- **Context everywhere** — TM, glossary (approved / forbidden terms) and style guidelines injected at every stage and optional RAG-based — domain adaptation *without* fine-tuning
- **Backend-agnostic** — plug in any NMT or LLM; deploy on new domains / languages immediately

WMT24++ — 11 languages, no TM / terminology

System	JA	ZH	DE	FR	ES	IT	IN	NO	VI	SW	PT
AIDA MultiLLM	27.9	47.2	31.5	**40.4**	**46.1**	43.6	34.8	45.1	37.0	41.5	42.0
AIDA LLM	29.6	41.5	33.7	40.1	44.5	**44.1**	35.9	45.5	37.1	43.6	41.7
AIDA NMT	**30.2**	**48.1**	**34.7**	38.7	44.8	43.2	**36.6**	**47.1**	**37.3**	42.8	**42.6**
GPT-4o	25.7	42.8	33.8	39.2	41.3	40.5	33.2	45.1	34.3	**44.9**	41.6
DeepL	26.3	39.2	33.4	38.2	45.5	42.9	32.5	44.1	–	44.1	41.3
Google Tr.	25.3	41.5	34.5	37.2	43.0	41.5	33.5	43.5	34.9	41.7	42.2
△ AIDA − best	**+3.9**	**+2.4**	**+0.2**	**+1.2**	**+0.6**	**+1.2**	**+3.1**	**+1.0**	**+2.4**	−1.3	**+0.4**

BLEU, EN→target (960 segments, 4 domains). AIDA Agents is **best on 10 of 11** languages — pure multi-agent orchestration, no retrieval. **Bold** = column best.

Industrial benchmark — full TM / terminology / style

	BLEU			Good/Exc. (%)		
Lang.	MS	Comp.	AIDA	MS	Comp.	AIDA
JA	24.0	52.8	**59.3**	12.2	85.7	69.4
ZH	24.9	43.1	**45.2**	62.2	82.7	77.6
FR	42.2	47.6	**48.8**	60.2	94.9	90.8
KO	48.1	54.2	**56.9**	73.4	82.9	81.6
DE	55.1	56.1	**57.0**	82.7	91.9	**98.0**
ES	58.5	64.4	**65.3**	92.9	95.9	94.9

MS = Microsoft NMT; Comp. = commercial system *with human post-editing*; AIDA = AIDA MultiLLM (*fully automated*). Highest BLEU on **all 6** languages.

Key findings

- **70–98%** of segments publication-ready with *no* human post-editing
- On German, the automated system **beats a human-post-edited competitor** (98.0% vs. 91.9%)
- Suggests a **50–70% reduction** in post-editing effort; human review remains fully integrable
- **Native XLIFF / CAT** integration ⇒ onboarding in hours, not weeks with direct reuse of existing linguistic resources.
- **Easy project-specific adaptation** ⇒ can be easily adapted to specific/non-standard project requirements

Conclusion

Unlike standalone NMT engines or research multi-agent systems, AIDA Agents is a **deployed orchestration platform** with native CAT/TMS integration and **no fine-tuning**. It incorporates any NMT or LLM backend, enables immediate deployment on new domains and languages, outperforms state-of-the-art MT on 11 WMT24++ languages, and delivers 70–98% publication-ready output industrially — even surpassing human-post-edited baselines for several pairs.

A terminology-specialized architecture ($AIDA_{term}$) is presented separately at *ACL 2026 Industry Track*.