

ANNI Guard & MT — Abschlussprotokoll & Runbook (Stand: 2025-08-30, Europe/Vienna)

1) Ziel & Kontext

Stabile End-to-End-Pipeline für **TranceLation Engine 3.0 über ANNI**. Fokus: **Guard (8091)** vor **MT-Worker (8090)**, Wahrung von Platzhaltern/HTML/Zahlen/Emojis, schnelle Fehlerdiagnose und reproduzierbare Start-/Stop-Prozedur.

2) Systemübersicht

Komponenten - **MT-Worker** (`mt_worker.py`), Port **8090**, führt die eigentliche Übersetzung aus. - **Guard** (`mt_guard.py`), Port **8091**, übernimmt Maskierung, Validierung (Placeholders/HTML/Nummern), Pivot-Routing, Glossar/TM und Fehlernormalisierung. - **TM/Glossar**: `tm.csv`, `glossary.json` (optional, lokal beim Guard).

Haupt-Endpoints - Worker: `GET /health`, `POST /translate` - Guard: `GET /health`, `GET /meta`, `POST /admin/reload`, `POST /translate`

Schlüssel-Eigenschaften Guard (aktuelle Implementierung) - Maskierung & Restore von `{{...}}`, `{app}`, `<tags>`, **Emojis**, `@™©SM`, **#Hashtags** - Absatz-Erhalt, **Pivot via EN** satzweise, De-quoting rund um Marker, Deduplikation doppelter HTML-Tags - Invarianten-Checks: `ph_ok`, `num_ok`, `html_ok`, `paren_ok`, `len_ratio`, Sammelwert `ok` - **CORS aktiv** (für GUI/Browser-Aufrufe) - **Fehlertransparenz**: 502 mit `backend_body`, wenn der Worker fehlschlägt

3) Rollen & Rechner

- **ANNI-Laptop** (Self-host MT + Guard). Beispiel-IP: `192.168.0.65`.
- **APP-Laptop** (Client/GUI/Tests). Ruft den Guard auf ANNI an.

Wichtig: Energiesparen/Sperrbildschirm am ANNI-Laptop deaktivieren, damit 8090/8091 erreichbar bleiben.

4) Start/Stop (ohne Kommentare, in Reihenfolge)

Auf ANNI starten

```
cd "$HOME/trancelate-onprem"
uvicorn mt_worker:app --host 0.0.0.0 --port 8090
```

Neues Terminal (ANNI):

```
cd "$HOME/trancelate-onprem"
MT_BACKEND=http://127.0.0.1:8090/translate
ANNI_API_KEY=topsecret
uvicorn mt_guard:app --host 0.0.0.0 --port 8091
```

Stoppen Port 8090/8091

```
lsof -tiTCP:8090 -sTCP:LISTEN -nP | xargs -I{} kill -9 {}
lsof -tiTCP:8091 -sTCP:LISTEN -nP | xargs -I{} kill -9 {}
```

5) Health & Meta Checks

Worker direkt

```
curl -sS "http://127.0.0.1:8090/health"
```

Guard

```
curl -sS "http://127.0.0.1:8091/health"
curl -sS "http://127.0.0.1:8091/meta"
```

Erwartung: `ok:true`, `backend_alive:true`, `backend_url:http://127.0.0.1:8090/translate`.

6) Übersetzen (Guard)

Einfacher Test

```
curl -sS -H 'Content-Type: application/json' -H 'X-API-Key: topsecret'
-X POST 'http://127.0.0.1:8091/translate'
--data-raw '{"text":"Hallo 😊 - 16:00","source":"de","target":"en"}'
```

Volle Probe mit Platzhaltern/HTML/Emojis/Zeiten/Ranges

```
curl -sS -H 'Content-Type: application/json' -H 'X-API-Key: topsecret'
-X POST 'http://127.0.0.1:8091/translate'
--data-raw '{"text":"Nur heute 🧐: {{COUNT}} Plätze frei bei <strong>{app}</strong>® - gültig für 2 Tage! Angebot endet um 4 pm (16:00) - Version ISO 9001, Python 3. Preise: 1.234,56 € inkl. MwSt.; Zeitraum 1990-2014; Hotline <a href=\"https://example.com\">hier</a>. #bewusst #verkaufen #mybesthealth\n\nKann das überhaupt funktionieren? Du versuchst, den Kunden wirklich zu verstehen - nicht nur seine Worte, sondern seine Ängste und
```

Sehnsüchte. Und wenn heute nichts zu holen ist, bleibst du trotzdem. (™/©)
😊, "source": "de", "target": "en"}'

Antwort enthält `translated_text` und `checks` (prüfe `html_ok`, `ph_ok`, `num_ok`).

7) Unterstützte Sprachpaare (aktuell direkt)

- ↔ **EN** für: DE, FR, ES, IT, NL, DA, SV, NO, PT
- Weitere Paare via **Pivot EN** (Guard übernimmt automatisch satzweise Pivot)

8) Translation Memory & Glossar

Reload

```
curl -sS -X POST "http://127.0.0.1:8091/admin/reload"
```

Dateien - `tm.csv` Spalten: `source_lang`, `target_lang`, `source_text`, `target_text` -
`glossary.json` Felder: `never_translate`: ["Begriff1", "Begriff2", ...]

TM-Strategie - Exact-Match zuerst, dann Fuzzy (falls `rapidfuzz` vorhanden), mit Platzhalter-Mengenvergleich - TM wird bei Emojis im Satz bewusst **übersprungen** (sonst nimmt MT den Satz)

9) Invarianten & Normalisierung

Checks - `ph_ok`: alle `{{...}}` und `{...}` im Ziel vorhanden - `html_ok`: gleiche Tag-Signatur
Quelle↔Ziel - `num_ok`: Zahlen/Zeiten konsistent (z.B. 16:00, Bereiche 1990–2014, AM/PM) -
`len_ratio`: Längenverhältnis innerhalb Limits

Normalisierung - De-quoting um `<tags>`, `{{...}}`, `{...}`, Emojis, Hashtags, Symbole ®™©SM -
Deduplikation doppelter HTML-Tags - Doppelpunkt-Regeln: `16:00` bleibt, „a: b“ wird „a: b“

10) CORS & GUI

- Guard und Worker haben CORS aktiviert (Allow-Origin: `*`) für lokale GUI-Tests.
- Falls Browser-CORS-Fehler auf 8090: Worker neu starten mit aktivierter CORS-Middleware (in `mt_worker.py`).

11) Fehlersuche (Schnellreferenz)

502 Bad Gateway vom Guard 1. Worker läuft nicht oder Port blockiert → Worker starten (8090) 2. Falscher `MT_BACKEND` → korrigieren (`http://127.0.0.1:8090/translate`) 3. Netzwerk/Schlafmodus → Energiesparen deaktivieren

500 Internal Server Error Worker - Fehlende Python-Pakete → im ANNI-Env installieren: `transformers sentencepiece sacremoses` (ggf. `torch` CPU) - Syntaxfehler in Guard/Worker → Datei kompilieren: `python3 -m py_compile mt_guard.py`

Port-Konflikt - Freigeben: `lsof -tiTCP:PORT -sTCP:LISTEN -nP | xargs -I{} kill -9 {}`

CORS-Fehler - CORS im Worker/Guard aktivieren, Browser neu laden

12) Betriebsregeln

- **Ein Schritt pro Aktion** (keine `&&`, keine Inline-Kommentare in Befehlen)
- Beim Testen immer **Rechner/Port** benennen (ANNI vs. APP)
- Nach Suspend/Sleep immer Health prüfen

13) Nächste Schritte

1. **Qualität:** Feinjustierung De-quote/Tag-Kollaps für tricky Sätze mit verschachteltem HTML
2. **Sprachen:** Osteuropa (RU, PL, CS, RO, HU, BG, UK, TR, EL, FI) mit soliden TM-Seeds und Testsätzen (≥ 2 Absätze, Emojis, Markenzeichen)
3. **GUI:** `anni_gui.html` zur Status-Zentrale ausbauen (Health/Meta, Logs, Batch-Tests)
4. **Persistenz:** Logging in `logs/` sichtbar machen, einfache Metrikseite (`/metrics`)

14) Quick-Smoke-Skripte

Mehrsprachig DE→(EN,FR,ES,IT,TR,RU)

```
python3 - <<'PY'
import requests
u="http://127.0.0.1:8091/translate"; k="topsecret"
t="""Nur heute 🤖: {{COUNT}} Plätze frei bei <strong>{app}</strong>@ - gültig
für 2 Tage! Angebot endet um 4 pm (16:00) - Version ISO 9001, Python 3.
Preise: 1.234,56 € inkl. MwSt.; Zeitraum 1990-2014; Hotline <a href='https://
example.com'>hier</a>. #bewusst #verkaufen #mybesthealth\n\nKann das
überhaupt funktionieren? Du versuchst, den Kunden wirklich zu verstehen -
nicht nur seine Worte, sondern seine Ängste und Sehnsüchte. Und wenn heute
nichts zu holen ist, bleibst du trotzdem. (™/©) 😊"""
for tgt in ["en","fr","es","it","tr","ru"]:
    r=requests.post(u,headers={"Content-Type":"application/json","X-API-
```

```
Key":k},json={"text":t,"source":"de","target":tgt},timeout=90)
print(tgt, r.status_code)
print(r.text[:320].replace("\n"," ↵ "))
PY
```

15) Change-Log (heute)

- Worker & Guard stabilisiert; CORS aktiviert
- Fehlerweitergabe 502 mit Worker-Body
- De-quote erweitert (Tags, Platzhalter, Emojis, Hashtags, Symbole)
- Doppel-Tag-Kollaps; Colon-Normalisierung; satzweises Pivot via EN
- Anchor-Innertext-Rettung
- TM/Glossar Reload-Endpoint