Case Study L1-003

Semantic Hallucination: The "Phantom Component"

Object Substitution in LiDAR Systems (Coupler vs. Lens)

Cédric Stéphany — Technical Translation & AI Alignment Specialist

Case Study Metadata

Dataset ID:	L1-003
Category:	Semantic Integrity — Level 1
Focus:	Hallucination / Component Substitution
Model:	Generic NMT
Domain:	LiDAR / Opto-Electronics

1 The Context: Technical Equivalence

In hardware patent claims, specific components define the physical embodiment of the invention. A "Coupler" and a "Lens" are distinct optical elements with opposing functions.

- **Coupler (Coupleur):** Splits or combines light signals (signal routing).
- **Lens (Lentille):** Focuses or diverges light beams (beam shaping).

Substituting one for the other is not a "translation error"—it is a fabrication of technology that does not exist in the source document.

Key Concept

The "2x2" Constraint:
The term "2x2" acts as a strong semantic modifier.

- In fiber optics, a **"2x2 coupler"** is a standard industry component.
- There is no standard component known as a **"2x2 lens"** or "2x2 diverging lens."

The presence of "2x2" should rigidly lock the terminology to the fiber optic domain (couplers/splitters), preventing drift into free-space optics (lenses).

2 The Glitch: "Creative" Hardware Invention

In this LiDAR patent, the generic NMT model encountered the phrase "*2x2 coupler*." Instead of retrieving the literal translation, the model hallucinated a completely different component: "*lentille de divergence*" (diverging lens).

This suggests the model associated "optical system" + "beam" + "2x2" (matrix) with "divergence/spreading," and overwrote the source text with a plausible-sounding but factually wrong component.

Critical Issue

The Fatal Flaw:
If filed, this translation would render the patent claim invalid for **Lack of Enablement**. The claim describes a system using a "lens" where the diagram and description clearly show a "coupler." A competitor could argue that the French patent describes a physically impossible or non-functioning device.

3 The Alignment Challenge

3.1 The Translation Failure

Source (English)	AI Hallucination (Failure)	Golden Rewrite (Correct)
"...wherein the optical coupler comprises a **2x2 coupler**."	× **Object Hallucination:** "...lequel l'élément optique comprend une **lentille de divergence**." (Invention of Non-Existent Component)	✓ **Semantic Precision:** "...lequel l'élément optique comprend un **coupleur 2x2**." (Technical Equivalent)

Table 1: Component Substitution in LiDAR Claim

4 Alignment Methodology

To eliminate these hallucinations, we employed a **Term-Level Drift Detection** workflow in Label Studio.

Alignment Methodology

Annotation Process:

1. **Tagging the Glitch:** As shown in the dataset, the annotator highlights the target phrase "*lentille de divergence*" and tags it with `HALLUCINATION_IA`.
2. **Source Anchor:** The source term "*2x2 coupler*" is tagged as `COMPOSANT_TECH` (Technical Component).
3. **Negative Reinforcement:** The model is penalized for "creative" substitutions. We enforce a strict dictionary match:
    - If Source = "*Coupler*"
    - Target CANNOT contain "*Lentille*"

This forces the model to abandon "probabilistic guessing" in favor of "literal technical mapping" for hardware components.

5 Results & Impact

- **Fact-Checking:** Restored physical accuracy to the patent claim.
- **Consistency:** Aligned with the patent drawings (which depict a coupler, not a lens).
- **Safety:** Prevented a "fatal flaw" rejection that would have required costly substantive examination arguments to correct.

6 Related Case Studies

- **C1-003:** Structural Compliance — System vs. Method Claim Morphology (Hyper-Correction in LiDAR)
- **L1-001:** "1-Hot" Hallucination — Context blindness in digital logic
- **L1-002:** "U-Turn" Disambiguation — Traffic vs. Optics polysemy