AI and LLMs in car transport

how and how not to use AI

Tobias Carlén · Axess Logistics · ECG

AXESS LOGISTICS

Introduction

this talk is on where LLMs and machine learning work well — and where they don't - it will cover

- real examples from Axess Logistics
- why it went like this
- a working approach for applying AI to real-world problems

AXESS LOGISTICS

subject: car dimensions

we have a major data quality problem with car dimensions and weight:

- many orders have incomplete or incorrect data
- OEM data is often generic, or based on older versions of the car
- manual entry errors, wrong unit of measure, data in the wrong fields…. etc etc etc

AXESS LOGISTICS

Approach: let AI fix it

train a model on a large set of correct vehicles, let AI check and correct orders for us.

> we had about **a million rows** of (mostly) correct order data — we trained first an **LLM**, then a **machine learning model**.

AXESS LOGISTICS

miserable results

- data quality improved slightly — but AI made ***lots of errors***
- AI overwrote good data with bad
- AI results were far worse than an experienced human guesstimating from brand and model

why did this fail?

the models created false correlations. a human knows that color has no real correlation to weight — AI does not.

also, minor errors in the training creates strange decisions.

AXESS LOGISTICS

instead we built a tree model

at the top sits a model — say Volvo XC60
with SQL queries that catches even badly misnamed XC60s,
and fallback average values for all dimensions, if we have nothing better

- **rule 0** — universal fallback · any non-empty make · 1680 × 4600 × 1880 mm / 1800 kg · 1,636 corrections
  - **rule 785** — any Scania = heavy truck · 3,562 corrections
  - **rule 319** — Volvo XC60 · 1675 × 4673 × 1898 mm / 1992 kg · 11,572 corrections
    - **319.6** — XC60 (II) · 1641 × 4689 × 1902 mm / 1831 kg · 57 corrections
    - **319.8** — XC60 (II) AWD · 1641 × 4691 × 1902 mm / 2020 kg (same body, heavier) · 831 corrections
  - **rule 228** — Model Y · 6,806 corrections
  - … ~2,000 model rules

AXESS LOGISTICS

transport models & inheritance

- each rule defines a **transport model** — a range of variants close enough to make no difference from a transport perspective
- each rule **inherits** everything it can from above — only what differs is specified
- values must fall within **ranges** — outliers get replaced by the transport model average
- identification varies: a piece of the model code, part of the name, something in the VIN — whatever we could find to make a rule
- external sources as cross-checks: a plate number gives us registry data, which for most cars is fairly accurate

AXESS LOGISTICS

now LLMs started making sense

with the tree model in place, improving it — fine-tuning rules, extending the rule set — turned out to be something an LLM can do!

> LLM can see new models, find the specs and draft new rules.
> It can look at exiting rules, analyze them and compare to online datasources and similar cars in our database
> **New rules and updates can be tested, verified and quickly deployed**

in hours, precision and coverage of our ruleset improved dramatically!

AXESS
LOGISTICS

improvement loop

LLM focuses on a rule and searchs out reference information → Identifies likely improvement potential; modifies rule / creates new rules → **run against test dataset** → measure: corrections improved? regressions? → **accept** → into production production **reject** → back to LLM

the LLM never outputs dimensions directly — it works to improve the correction rules. A human checks all changes before production - Not all suggested changes are good!

same loop a human expert would run. minutes, not a training cycle.

AXESS LOGISTICS

the result

- after the correction engine runs, **every car** has data for every dimension
- **>99%** of them are exactly right
- on average **14%** of orders are corrected by the current rule set

numbers from shadow-prd snapshot, 2026-06-09

AXESS LOGISTICS

why it worked

- AI doesn't have common sense — you must establish **clear boundaries** and show where **true correlation** exists
- a strict system like the tree model is easily understood by an LLM — it can work within the rules *and* suggest changes to them
- a big **test data set** (cars with known real dimensions) gives quick, exact evaluation: *"did it improve predictions? did it destroy any data?"*

> finding and learning from mistakes is far more effective than trying to force the LLM *not* to make mistakes.

AXESS LOGISTICS

use LLMs to maintain the structure, not to be the structure

1. **define the boundaries** — with specific, hierarchical examples: a passenger car weighs 900–3500 kg; an average car ~1700 kg; a compact car 1200–1500 kg
2. **point out valid correlations** — drive train impacts weight significantly
3. **point out false correlations** — color has no correlation to weight
4. **define a system for AI to work within** — fine to let AI help build it, but nail the problem to a framework first
5. **create a test data set** — inputs and expected outputs, like any normal testing; fast discovery when something goes wrong
6. **document all assumptions carefully** — let AI do this for you! when something goes bad, check assumptions first

AXESS LOGISTICS

what LLMs do — and don't do — well

good at

- generic problems with a large collection of similar problems already solved (e.g. programming)
- adjusting parameters, like in our example
- small adjustments to existing models
- reading large amounts of data, finding patterns and mistakes

not good at

- deterministic relationships (axle weight in a truck) — when the outcome can be exactly calculated, use a deterministic model
- complex optimization — use an optimization engine
- anything outside the training scope — answers turn to garbage
- being the last line — LLMs fail catastrophically without warning; other systems or people must step in

AXESS LOGISTICS

what this means for human skills

- Real experts become even more necessary - people who really understand how your business works are vital to implement AI
- System thinkers are also vital - people who can extract knowledge from experts and codify it, turn it into useable rules that systems, including AI, can understand
- Some people MAY be able to do both, but it's rare, and this is still something that needs to be done collaboratively; it turns out experts do things very differently, and it takes discussion and analysis to determine best practise

AXESS LOGISTICS

leverage from this - an example

- Your best planner will only be able to plan a limited number of trucks and orders every day
- if all your best planners distill their best thinking and decision rules for doing things right, you can have a model that combines them and applies them tirelessly and rigorously, to ALL trucks and ALL orders
- Planners can shift their focus from planning trucks to tuning the rules, handling special cases as they come up and thinking about how to handle similar cases automatically

AI is a great tool — but it's just part of the right toolkit

LLMs are trained on large volumes of text. they are very good at interpreting language. they are **not solution engines** — design the structure of the solution first, then AI can feed and adjust it.

combine AI with deterministic models, SQL, optimization engines and **human common sense** to produce the best results.

AXESS LOGISTICS

thanks for listening!

questions, ideas, war stories:

tobias.carlen@axesslogistics.com

tobcar.se

(link to online version of this presentation)

AXESS LOGISTICS