Agentic AI use at Belle II

23/10/2026

Huw Haigh

Where does Belle II/SuperKEKB stand in regards to agentic AI/LLM use?

Belle II AI Tools Policy

- Task force formed by EB in May 2026, directed at defining a policy document.
- First draft shared this week from B2GM - focus mainly on guidelines/ethics of data use (personal data, experimental data).
- Main discussion around **what exactly constitutes "use of unpublished data"**, correct usage seems rather difficult to clearly outline - initial definition means many users have probably already broken this…

Development of LLM framework

- SuperKEKB AI assistant framework established. Has access to huge database of beam data.
- Conversational interface allowing quick beam diagnostics - e.g was prompted "Investigate beam emittance variation over time", discovered notable correlation with part of vertical steering magnets and directed optics experts immediately to that location.
- Discussions are still ongoing with Belle II LLM implementation, unclear if it is actually wanted.

What have collaborators been saying about integrating AI/LLMs into Belle II analysis flows

Positive? **Neutral?** **Negative?**

A chatbot (with Belle II internal knowledge) would be very useful.

Environmental impact has not been considered - current AI tools are unsustainable.

Current AI tools are very good at producing **convincing but wrong** answers.

Users still must know how to code, otherwise they risk missing AI-induced mistakes.

API tools to include in existing tools (BASF2?) could would be useful.

The large resources required to develop Belle II specific AI (with understanding of Belle Analysis Framework) would be **far better spent on proper documentation** and manual bug fixing.

The idea of having to pass a review committee with analysis written by AI is terrifying - I can see nothing but delays.

Agentic AI/LLM use of Belle II members at MBI covers full spectrum:

No use at all

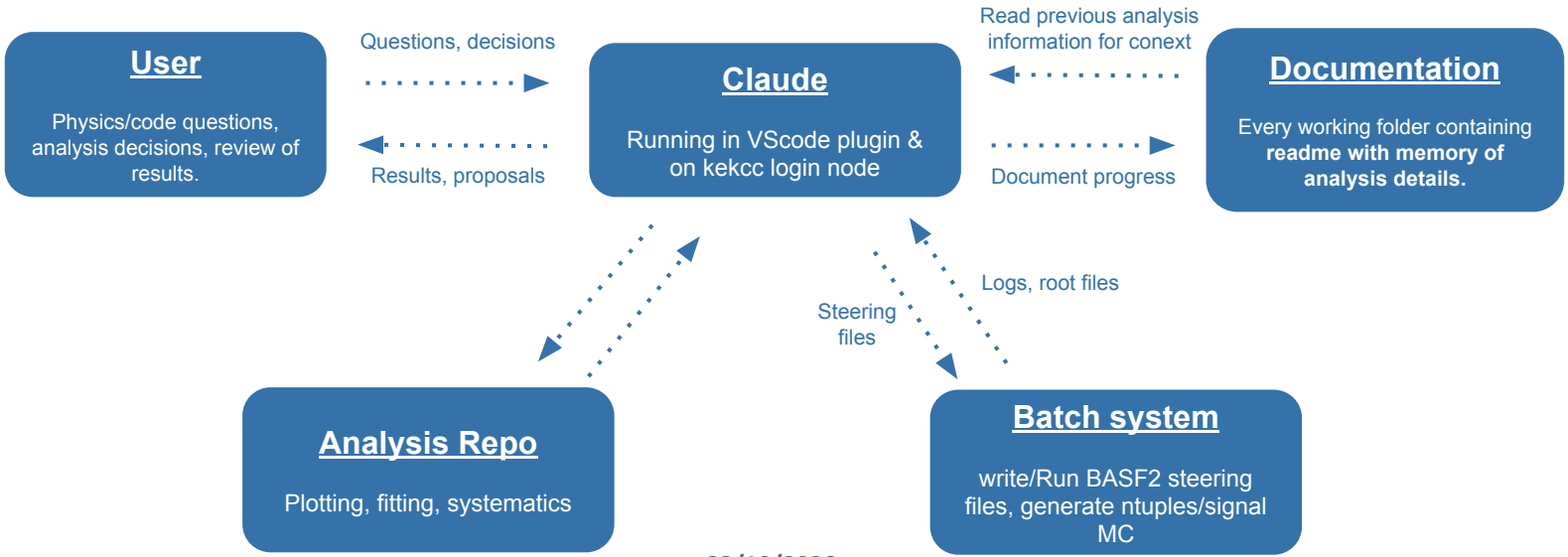
Brings no advantage - code already does everything needed for Belle II physics analysis

Medium usage

Individual complex analysis tasks (i.e. plotting, uncertainty propagation), greatly speeds up completion

High usage

Agent is conducting full analysis from initial data processing to plotting, presentation.



How LLMs fit into a full analysis workflow at Belle II.

An example - I want to scope out a possible physics analysis, using inclusive tagging approach to look for $B \to \omega \nu\nu$ decays.

1. Brainstorming

Claude: Discusses feasibility of analysis and main considerations (i.e. points out π^0 reconstruction problems), **theory background, literature review**.

2. Reconstruction code

Claude: Decides signal signature in detector ($\omega \to \pi^0\pi^+\pi^-$), **provides relatively simple $B \to \omega\nu\nu$ steering file**, decay.dec files

3. MVA event selection

Claude: Suggests methods, based on literature review (i.e. **suggested 2 stage MVA, 1 looks for good ω decay, second focuses on event shape.**), runs complex tests to optimise selection.

4. Validation + systematics

Claude: Suggests **control channels (i.e. $B \to \omega J/\psi$ for π^0 reco. validation)**, systematics and methods to measure them, **writes production and analysis code**

5. Signal Extraction

Claude: Suggests signal extraction methods (i.e. **3D pyhf fit, classifier output x $m(\pi^0\pi^+\pi^-)$ x ROE_extraE**), implements systematics from validation, writes/tests fitting code.

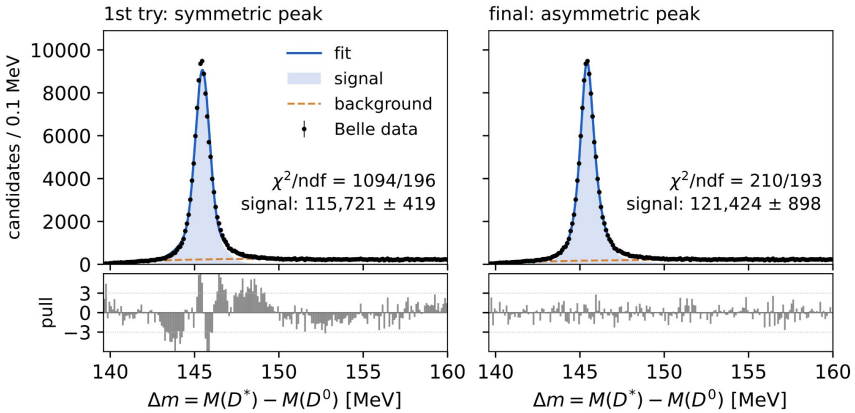
5. Writing

Claude: Review of written work, can be useful in Latex formatting (I would use far more sparingly for writing tasks)

- **Task:** π/K ⟶μ particleID fakerates in data and MC, needed for lepton flavour universality study.
- **Prompt:** Find or generate a D^* control sample with decay $D^{*+} \rightarrow D^0(\rightarrow K^-\pi^+)\pi^+_{slow}$, derive particle fakerates by way of fit the ΔM peak.

- **Agents actions:**
  - Finds required data samples on KEKCC, writes/runs reconstruction code.
  - Writes fit code (RooFit: gaussian + polynomial background) -> run fit, check results.
  - **Agent notices pulls show peak has tail not fully captured with symmetric Gaussian** -> modifies function to contain power tail, verifies improvement.
  - Plots full results to user, summarise work and findings in documentation readme

=> **total analysis time take just minutes.**

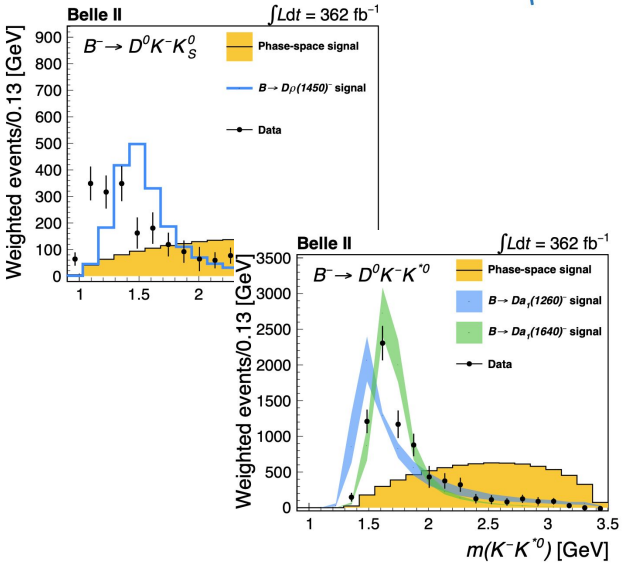


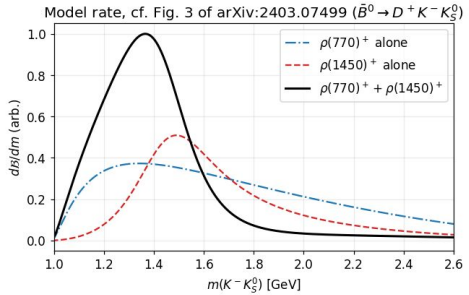
- **Task:** B->DKK is a small background in B^+->$K^+\nu\nu$ search at Belle. Belle II study found this is mismodelled in MC, **standard phase-space assumption doesn't hold as decay actually proceeds via multiple resonances**. Need to reweight by hand.
- **Prompt:** Here are **two papers** outlining theory and a search for the B->DKK. We need to **produce and reweight a dedicated MC sample**, define associated systematics and include it in the binned fit.

arXiv:2406.06277 arXiv:2403.07499

- **Agents actions:**
  - Analyses both papers to understand task at hand, report back summary of understanding + clarification questions.
  - Writes MC simulation/production code for dedicated sample, **complex scripts to implement reweighting** + systematics propagation
  - **Found and diagnosed a 10-12% bias in initial reweighting** due to initial assumptions - updated code to account for this.
  - Produced fully processed ntuples, **updated generic MC samples by removing PSHP B->DKK, implemented new dedicate sample across full analysis repo.**

=> **Reduced a multi-day job down to less than an hour of work**





Pro

- **Massively speeds up coding, can write very complex code quickly.**
- **Quick access to great depth of information - can find relevant papers/studies.**
- Explains and derives claims, can quickly produce dedicated studies to prove assumptions.
- **Can handle large scale process, batch jobs quicker and reliably**
- Can find silent failures human may miss.

Con

- **Overuse leads to removal of humans from scientific process - could actually slow things down**
- **Often makes confident but wrong statements.**
- Doesn't always have access to important contextual information (i.e. secure documentation)
- Operational slips - can interfere with other works from user.
- **Important context information can be quietly forgotten.**
- Mission creep (initial idea/task gets lost in rabbit hole of code slop), lose track of codebase
- **Negative impacts on collaborative aspects.**
- **Environmental concerns**

Agentic AI is obviously an exceedingly powerful tool, and most (but not all) Belle II members use it regularly. It can easily be incorporated into **every level of the analysis flow**, particularly:

- Implementation of complex analysis requirements
- Maintenance of large code repositories
- Efficient batch job handling
- Literature review, planning
- Code/writing review

However, **everyone expresses some level of concern in one way or another**, ranging from environmental cost to the dehumanisation of our writing and review processes.

I believe we must **spare a thought for how this will impact the education/development of the next generation of physicists** - we must take care to not lose the very expertise that these AIs have been trained on.

> "In some areas, particularly in education and in the training of young mathematicians, it will be **crucial to emphasize the irreducibly human aspect of our work**, and to restrict the use of AI tools quite tightly; **the goal of training a mathematician is not achieved by producing correct homework.** "
>
> - *Mathematics in the age of AI*, Terence Tao