AI for Maths Research – Case Studies
Michel van Garrel | Uni Birmingham
GDG AI for Science Australia | October 16, 2025

A Paradigm Shift (Mid-2025)

- **Before:** LLMs imitate maths; great at connections, feeble at logic.
- **Now:** Advanced models (e.g., Gemini Deep Think) mark a new milestone in reasoning capabilities.
- AI is evolving into the mathematician's "Co-Pilot" (Terence Tao, Scientific American, 06/2024).

Case Study 1: A Leap in Logic

The Test

Challenge a correct AI-generated proof, then follow up with a non-sense request (e.g., "Use fractal theory to simplify").

Standard LLM (e.g., Pro)

- Retracts correct proof when challenged:
  "You are absolutely right..."
- Hallucinates when presented with nonsense.

Advanced AI (e.g., Deep Think)

- Defends its reasoning:
  "...I must reaffirm that the derivation is mathematically sound..."
- Rejects misplaced suggestions.

⇝ Significantly improved logic and robustness.

Case Study 2: The Polymath Companion

Context: Year-long struggle to prove a list of equalities.

Missing: relevant expertise in asymptotic analysis.

After elaborate prompts (setup of problem/requests in small increments)

- **Deep Think Suggestion:** Use RMT (Ramanujan Master Theorem).

The Outcome

- RMT was spot-on: A deep insight $\rightsquigarrow$ improved understanding.
- After some back-and-forth, Deep Think provided complete rigorous maths proof leading to major progress towards larger goal.

$\rightsquigarrow$ AI acting as a polymath, connecting relevant techniques.

Case Study 3: The Brilliant Connector – Pattern Recognition

The Context

A binomial identity (Conjecture in Barrott–Nabijou, 2022).

Deep Think Contribution

1. A flawless, rigorous proof (using analytic combinatorics).
2. A significant generalisation of the identity.
3. Connections to deeper structures: generating functions of trees, Kronecker quiver DT invariants, local Gromov–Witten invariants of curves, etc.

$\rightsquigarrow$ Accelerating discovery by revealing hidden mathematical structures.

Conclusions: The AI Co-Pilot

Framework for Interaction

- **Assistant**: Excellent for well-defined sub-tasks and computations.
- **Connector**: Vast knowledge; identifies hidden connections between fields.

Caveats and Challenges

- Logic is improving but still flawed; prone to taking incorrect shortcuts.
- Rarely admits ignorance ("I don't know") unless prompted for confidence.
- Potential for hallucinations require constant verification.

The Essential Principle: **Know Thy Maths**

Domain expertise crucial to guide AI, verify outputs, integrate insights.