Workshop Evaluation Summary

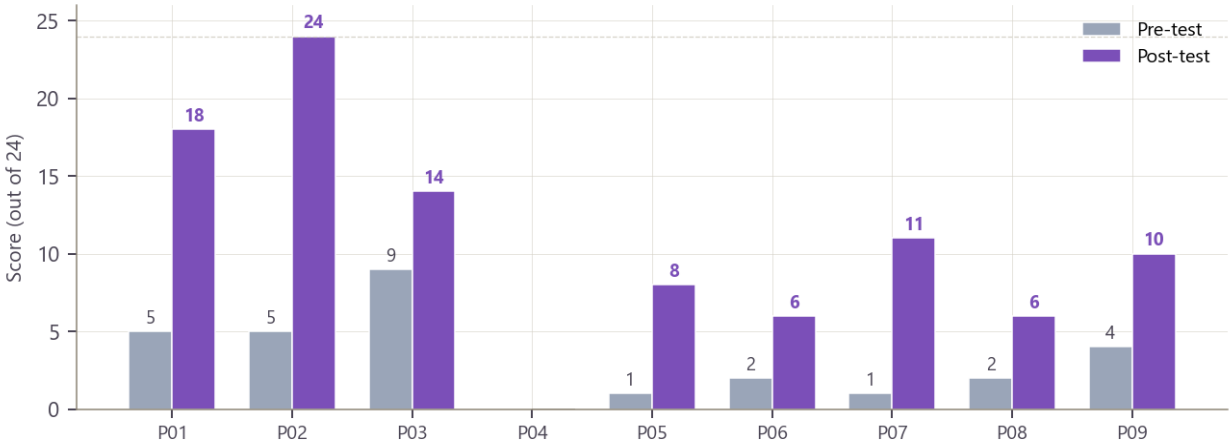
Bridging computational foundations to generative AI

9 participants · 6 modules · February 2026

Dear participant — this brief report summarises the data collected during the workshop. A 24-item knowledge assessment was administered before and after the day; cognitive load was sampled after every module using the NASA Task Load Index; exit tickets recorded self-rated understanding and open reflection. Numbers are rounded to one decimal place and effect sizes are reported as exploratory pilot-stage estimates.

§ I

Group result



Pre-test (left) and post-test (right) scores out of 24, by participant. P04 did not complete the post-test (hatched bar).

COHEN'S D	MEAN GAIN	NORM. GAIN
1.615	**+8.5**	**g = 0.43**
Very large effect (pilot, n=8)	*out of 24 points · 3.6 → 12.1*	*Hake's normalised gain · 'medium'*

Every participant who completed both tests improved; none regressed. The Wilcoxon signed-rank test confirms the shift is statistically detectable even at the pilot sample size (p = 0.008). Per-section analysis below shows the gain concentrated in Sections A and C — Section D (generative AI) moved more modestly, consistent with its placement at the close of a long day.

§ II

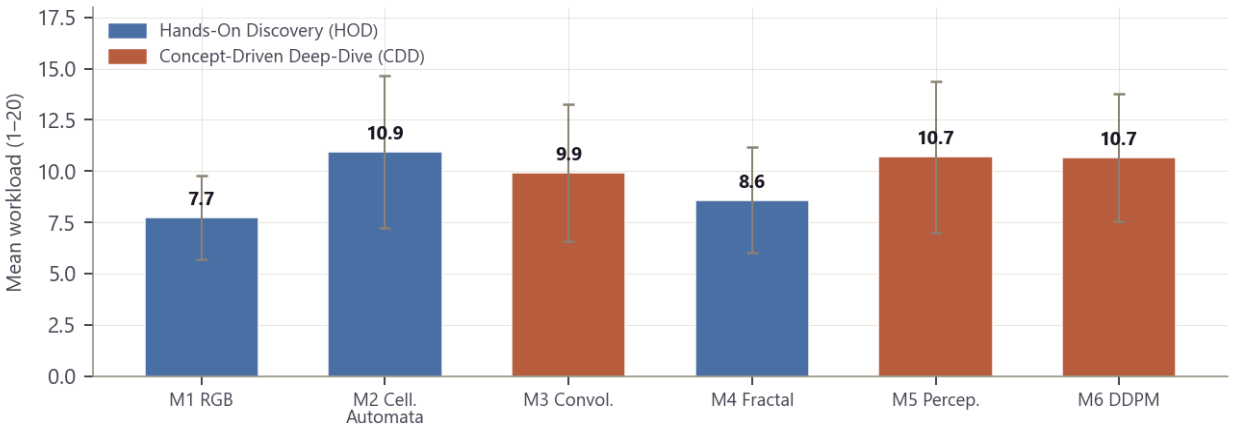
Individual scores

Participant	Pre-test	Post-test	Gain	Category
P01	5 / 24	18 / 24	+13	**Medium**
P02	5 / 24	24 / 24	+19	**High**
P03	9 / 24	14 / 24	+5	**Medium**
P04	*0 / 24*	—	—	—
P05	1 / 24	8 / 24	+7	**Medium**
P06	2 / 24	6 / 24	+4	**Low**
P07	1 / 24	11 / 24	+10	**Medium**
P08	2 / 24	6 / 24	+4	**Low**
P09	4 / 24	10 / 24	+6	**Medium**

P04 did not complete the post-test and is excluded from gain calculations. Categories use Hake's normalised-gain bands: High (g ≥ 0.7), Medium (0.3–0.7), Low (< 0.3).

§ III

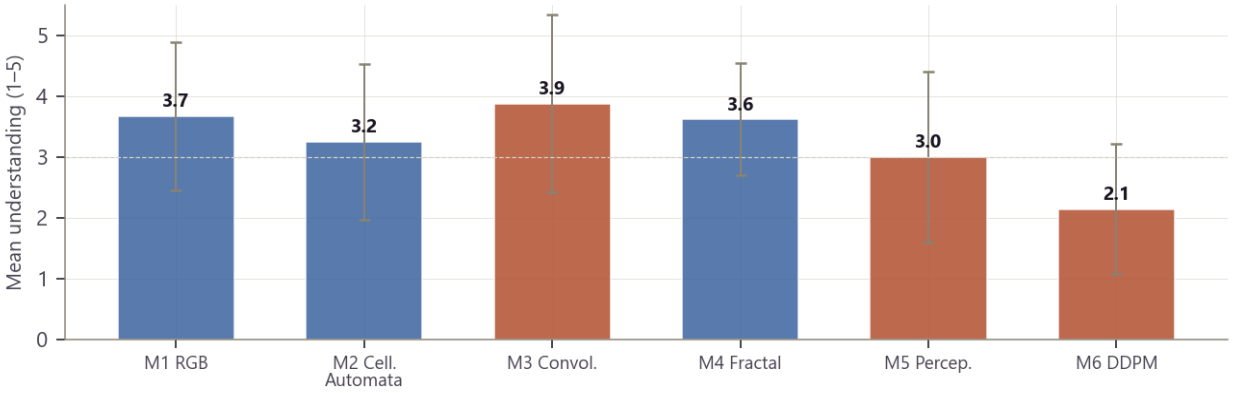
Cognitive load by module



NASA-TLX overall workload averaged across participants for each of the six modules. Error bars show ±1 standard deviation. The trajectory is an inverted-U peaking at the cellular-automata and perceptron lessons; load returns to moderate values by the closing module.

§ IV

Self-rated understanding



Exit-ticket Q4 responses on a 1–5 scale ('How well do you feel you understood this module?'). The dashed line marks the scale midpoint. Self-rated understanding tracks the inverse of cognitive load with one notable exception: M3 Convolution, which scored highly on both load and understanding.

§ V

Framework comparison

MEAN WORKLOAD (TLX)	SELF-RATED UNDERSTANDING	DIRECTION
9.1 / 10.4	***3.5 / 3.0***	***HOD lower***
HOD modules vs CDD modules · scale 1–20	*HOD vs CDD · scale 1–5*	*across both measures; differences are descriptive*

The workshop alternated between two pedagogical approaches. **Hands-On Discovery** (Modules 1, 2, 4) opens with code and visual feedback, introducing concepts as they become necessary. **Concept-Driven Deep-Dive** (Modules 3, 5, 6) opens with theory and guides participants into implementation. On average HOD modules were rated as slightly less demanding and received slightly higher self-rated understanding scores; the difference was largest on the mental-demand sub-scale (CDD 12.7 vs HOD 9.7 out of 20).

An important caveat: CDD modules covered inherently more complex topics (convolution, perceptrons, diffusion). Some of the load differential almost certainly reflects topic difficulty rather than pedagogical approach alone. Disentangling the two would require a controlled experiment that the pilot could not run.

§ VI

Key takeaways

1. Every participant who completed both administrations improved on the conceptual assessment, with an average normalised gain in the 'medium' band — meaningful learning across the group given the single-weekend format.

2. The strongest gains were in Section A (Arrays & Images) and Section C (Neural Networks). Section D (Generative AI) was the most challenging, consistent with the conceptual density of diffusion models and the placement at the close of the day.

3. Cognitive load increased with module complexity but remained manageable — every module's mean workload stayed below the midpoint of the 1–20 NASA-TLX scale.

4. Prior programming experience predicted lower reported load (Spearman $\rho = -0.857$, $p = .006$). The curriculum is, at present, calibrated for the less-experienced end of the participant distribution; differentiation for more-experienced learners is a target for the next design cycle.

§ VII

Limitations

The findings are pilot-stage. The sample ($n = 9$; $n = 8$ for paired pre/post) is small by inferential-statistics standards, so effect sizes carry wide confidence intervals. The workshop was a single contiguous session, so the gains may reflect that intensity and not generalise to a distributed schedule. Transfer was sampled via stated intention only; delayed re-administration was not part of this cycle. The thesis discusses each constraint in detail and proposes empirical follow-up for the next cycle of the curriculum.

Thank you for your participation and contribution to this research.