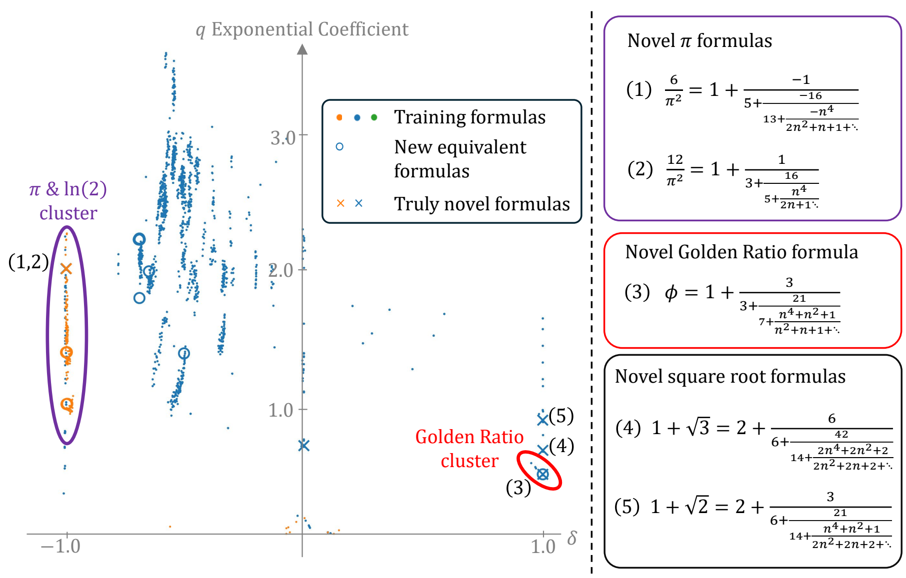
Ramanujan Machine: Unsupervised Formula Conjecture Generation

Michael Shalyt, Uri Seligmann, Rotem Elimelech, Shachar Weinbaum, Rotem Kalisch, Tomer Raz, Hila Barkan, Itay Beit Halachmi, Sheve Leibtag, Ido Kaminer.

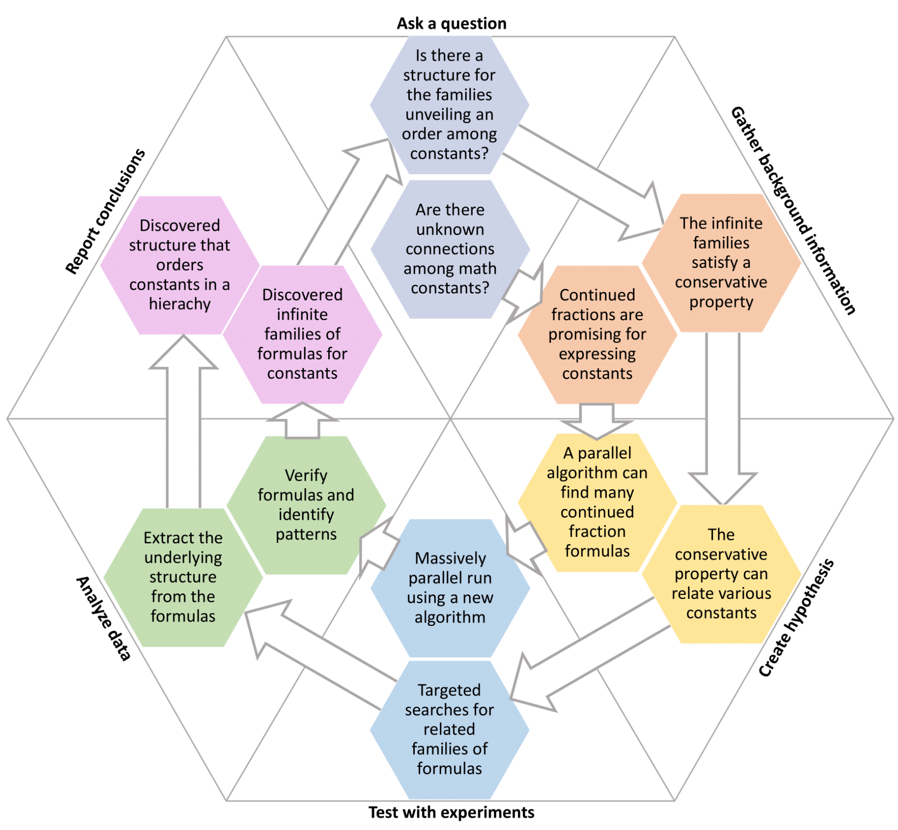
Ongoing efforts that span over decades show a rise of automated methods for scientific discovery and hypothesis creation [Fajtlowicz, 1988, Petkovsek et al., 1996, Wolfram et al., 2002, Buchberger et al., 2006, Bailey et al., 2007, Raayoni et al., 2021, Davies et al., 2021, Fawzi et al., 2022]. Despite the significant advances in Automated Theorem Provers, Automated Conjecture Generation in mathematics, especially in the field of number theory, remains a persistent challenge for AI - even for cutting edge LLMs. Specifically, AI methods were not effective in creation of formulas for mathematical constants because such formulas are either true or false - with no "distance metric" between two formulas, preventing continuous adjustments that can enhance their correctness. The absence of a systematic method left the realm of formula discovery elusive for automated methods.

This talk will present our recent demonstration of a systematic methodology for categorization, characterization, and pattern identification of such formulas. The key to our methodology is introducing metrics based on the convergence dynamics of the formulas, which we utilize for the first automated clustering of mathematical formulas.

The methodology was tested on a set of 1,768,900 such formulas, identifying many known formulas for mathematical constants, and discovering previously unknown formulas for π, π^2, $\ln(2)$, Gauss, and Lemniscate constants. The uncovered patterns enable a direct generalization of individual formulas to infinite families, unveiling rich mathematical structures.



New formulas are discovered and labeled based on clusters trained on simpler formulas, showing generalization. Horizontal axis represents the irrationality measure δ, vertical axis represents the exponential component of the approximation's denominator growth rate as a function of depth n.



Our talk will also review other recent contributions by the Ramanujan Machine group [Elimelech et al.,2024]. We developed a massively parallel algorithm that found an unprecedented number of continued fraction formulas for numerous mathematical constants. The patterns arising from those formulas enabled the first computer-assisted discovery of a mathematical structure: Conservative Matrix Fields, hinting at an order among mathematical constants and exemplifying the power of hybrid human-machine research (see cycle of discovery on the left).

We will further discuss generalizations of our proof-of-concept works in AI for mathematics. Specifically using modern approaches from ML in new ways and on different challenges, generating hypothesis via sifting instead of direct construction, using a small set of known examples to label clusters and generalize into higher complexity structures.