

ВИКОРИСТАННЯ AI В АВТОМАТИЧНОМУ ДОВЕДЕННІ МАТЕМАТИЧНИХ ТВЕРДЖЕНЬ

Вінницький національний технічний університет

Анотація

В роботі досліджено застосування ШІ в автоматичному доведенні математичних тверджень, детально розглянуто одну з найновіших та найуспішніших моделей - Goedel-Prover, розглянуто системи на основі GPT та Isabelle/HOL (створену у Кембріджському університеті). Проаналізовано методи формалізації та генерування доведень, а також представлення та побудову доведень математичних знань у цих системах. Особлива увага приділяється моделі Goedel-Prover як прикладу успішної інтеграції великих мовних моделей із можливостями формальної верифікації мови доведення теорем Lean 4. Через порівняльний аналіз визначено сильні сторони та обмеження кожного підходу, підкреслено переваги, недоліки та можливості для подальших досліджень у сфері математичного міркування за допомогою ШІ.

Ключові слова: Isabelle, штучний інтелект, логіка, GPT, мовні моделі.

Abstract

This paper examines the application of AI in automating mathematical proofs, with particular focus on three significant approaches: the Goedel-Prover model, Isabelle/HOL (developed by University of Cambridge) and, of course, GPT-based systems. We analyze the methodologies employed for formalizing and generating proofs, as well as the representation and construction of mathematical knowledge within these systems. Special attention is given to the Goedel-Prover model as a case study of successful integration between Large Language Models and the formal verification capabilities of the Lean 4 theorem proving language. Through comparative analysis, we identify the strengths and limitations of each approach, including advantages, disadvantages and opportunities for future research in AI-assisted mathematical reasoning.

Keywords: Isabelle, artificial intelligence, logic, GPT, language models.

Вступ

Останніми роками зростання інтересу до штучного інтелекту суттєво вплинуло на підходи до розв'язання математичних задач. Один із перспективних напрямів — автоматизоване доведення математичних теорем із використанням ШІ.

Мета роботи — розглянути, як саме штучний інтелект допомагає у здійсненні математичних доведень за допомогою моделі Goedel-Prover, системі Isabelle/HOL (розробленій у Кембріджському університеті) та мовним моделям GPT; проаналізувати, як ці системи формалізують знання та будують доведення, з особливим акцентом на Isabelle як приклад вдалої інтеграції великих мовних моделей із формальними інструментами, які надає мова Standard ML[3].

Результати дослідження

Механізм роботи системи Isabelle:

1.1 Формалізація твердження: в Isabelle користувач формулює задачу в термінах конкретної логіки, наприклад, логіки вищого порядку (HOL). Це означає, що для того, щоб довести якесь математичне твердження, користувач має чітко визначити функції, аксіоми, які будуть використовуватися у доведенні. Кожен елемент задачі має бути чітко описаний і математично коректно сформульований. Наприклад, можна визначити аксіоми, на яких буде ґрунтуватися доказ чи встановити зв'язки між різними математичними об'єктами.

1.2 Розбиття задачі на підзадачі: якщо твердження складне, Isabelle дозволяє розбивати доказ на менші підзадачі. Це необхідно для того, щоб кожен крок доведення можна було б розглядати окремо. Користувач може застосовувати тактики — інструменти, що автоматично чи частково виконують доведення певної частини задачі. Кожна підзадача обробляється окремо, що дозволяє більш ефективно знаходити рішення складних питань. Часто для доведення кожної підзадачі потрібно застосувати специфічні стратегії або тактики, в залежності від її складності.

1.3 Автоматичні та ручні докази: Isabelle забезпечує інтеграцію з різними автоматизованими засобами доведення, такими як SMT-солвери (наприклад, Z3) та автоматичні теоремні перевірки (ATP) [5]. Це дозволяє значно прискорити процес доведення через автоматичне вирішення стандартних підзадач. Однак складніші або нестандартні частини доведення часто вимагають втручання користувача. У таких випадках користувач може вручну застосовувати тактики для вирішення проблеми.

1.4 Формальна перевірка ядром: однією з головних характеристик Isabelle є її ядро, яке перевіряє кожен крок доведення на коректність. Це забезпечує високу надійність і гарантію того, що кожен доведений результат є математично бездоганим. Ядро Isabelle базується на системі LCF (Logic for Computable Functions), яка слідкує за тим, щоб кожен крок доведення був виведений за допомогою коректних логічних правил [1].

1.5 Побудова фінального доказу: коли всі підзадачі доведені, Isabelle автоматично складає фінальний доказ. Цей доказ є формальним підтвердженням істинності початкового твердження. Завдяки формальному характеру доказу, користувач може бути впевнений у його коректності.

Переваги використання Isabelle:

2.1 Формальна надійність: Isabelle дає можливість досягти формальної надійності, оскільки кожен крок доведення перевіряється на коректність ядром системи. Таким чином, кожне твердження, яке було доведено, є математично бездоганим. Це є особливо важливим у контексті формальних наук, де навіть одна помилка в доказі може призвести до серйозних наслідків.

2.2 Гнучкість логік: Isabelle підтримує не лише логіку вищого порядку (HOL), але й інші логічні системи, що дає можливість адаптувати систему до різних типів задач. Це дозволяє Isabelle бути універсальним інструментом, який можна використовувати для формалізації теорій в різних галузях математики та комп'ютерних наук.

2.3 Автоматизація рутинних частин доказу: Isabelle забезпечує часткову автоматизацію рутинних операцій у процесі доведення. Наприклад, автоматичне застосування стандартних правил чи теорем, спрощення виразів, перевірка на контрприклад — все це можна автоматизувати за допомогою інтегрованих інструментів. Це дозволяє значно зменшити обсяг роботи для користувача, але в той же час залишає простір для ручного втручання, коли потрібно.

2.4 Модульність: іншою важливою перевагою Isabelle є її модульність. Користувачі можуть використовувати вже існуючі доведення як будівельні блоки для нових доказів. Це дозволяє швидко та ефективно будувати складні докази, використовуючи вже доведені теореми чи підсистеми.

2.5 Велика база формалізованих знань: Isabelle містить велику базу формалізованих знань, зокрема в Isabelle Archive of Formal Proofs (AFP), що включає велику кількість теорем, доведених за допомогою Isabelle. Це дає користувачам змогу використовувати вже готові доказові матеріали та на основі їх будувати нові теореми. Така база є дуже корисною, оскільки знижує необхідність знову доводити те, що вже було доведено [4].

Недоліками Isabelle є складність освоєння, що є одним із найбільших недоліків. Для ефективного використання системи необхідно мати глибокі знання формальної логіки та досвід роботи з системами доведення. Користувачам без математичної підготовки може бути складно зрозуміти основи, що може ускладнити перші етапи навчання. Також, часткова автоматизація є однією з поганих рис хоча Isabelle підтримує автоматизовані інструменти доведення, не всі частини доведення можуть бути автоматизовані. Складніші, нестандартні чи спеціалізовані задачі часто вимагають втручання користувача для визначення шляхів доведення або вибору конкретних тактик. Останнім недоліком є те, що процес доведення в Isabelle може бути дуже ресурсоємним, особливо коли мова йде про великі

теореми або складні програми. Для таких задач може знадобитися потужне апаратне забезпечення, оскільки для їх вирішення потрібні значні обчислювальні ресурси [2].

Відмінності між Isabelle та мовними моделями GPT полягають в принципі роботи, автоматизації та області застосування. Isabelle та мовні моделі, такі як GPT, працюють за принципово різними механізмами, що відображає їх різне призначення та область застосування.

Isabelle працює на основі формальних логічних правил, що гарантує математичну точність і коректність кожного кроку доведення. Вона не допускає жодних помилок в доведенні, оскільки кожен крок перевіряється ядром системи. Це дозволяє забезпечити строгість і надійність доведень. Натомість GPT базується на статистичних моделях, які генерують текст, спираючись на ймовірності і залежності, що існують у великих наборах текстових даних. GPT не гарантує істинність своїх відповідей, і хоча вона здатна генерувати відповіді на основі контексту, ці відповіді можуть бути недостовірними або некоректними. Isabelle автоматизує лише певні частини процесу доведення. Наприклад, застосування стандартних тактик чи інтеграція зовнішніх солверів для певних завдань. Однак складні або спеціалізовані аспекти вимагають втручання людини. GPT ж є повністю автоматизованою системою, що генерує відповіді без участі користувача. Однак він не перевіряє коректність своїх відповідей, що може призводити до помилок чи неточностей.

Isabelle спеціалізується на формалізації математики та верифікації програм, забезпечуючи строгість і точність у доведеннях. GPT, у свою чергу, використовується для генерації природномовних відповідей, що підходить для завдань на зразок спілкування, творчих завдань чи автоматичних перекладів.

Висновки

Автоматизоване доведення математичних тверджень за допомогою штучного інтелекту відкриває нові перспективи у розвитку математичних досліджень. Isabelle є незамінним інструментом для формальних доказів та перевірок, забезпечуючи точність і надійність завдяки строгим логічним правилам. Мовні моделі GPT, натомість, орієнтовані на генерацію тексту та обробку природної мови, проте не забезпечують такої ж точності та гарантії істинності. Інтеграція сучасних мовних моделей із формальними довідниковими системами, як у випадку Goedel-Prover, показує перспективний напрям розвитку, поєднуючи сильні сторони обох підходів. У результаті порівняння виявлено переваги та недоліки цього підходу, а також окреслено потенційні напрями подальших досліджень у сфері математичного мислення за допомогою ШІ. У подальшому дослідження в цій галузі мають бути спрямовані на підвищення рівня автоматизації формальних доказів без втрати математичної точності.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Isabelle (proof assistant). [Електронний ресурс] – Режим доступу: [https://en.m.wikipedia.org/wiki/Isabelle_\(proof_assistant\)](https://en.m.wikipedia.org/wiki/Isabelle_(proof_assistant)) (дата звернення: 29.04.2025).
2. The Isabelle/Isar Reference Manual. [Електронний ресурс] – Режим доступу: <https://isabelle.in.tum.de/doc/isar-ref.pdf> (дата звернення: 29.04.2025).
3. [Електронний ресурс] – Режим доступу: <https://isabelle.in.tum.de/overview.html> (дата звернення: 29.04.2025).
4. Isabelle/jEdit as IDE for domain-specific formal languages and informal text documents. [Електронний ресурс] – Режим доступу: <https://sketis.net/wp-content/uploads/2018/05/isabelle-jedit-fide2018.pdf> (дата звернення: 29.04.2025).
5. Proof Reconstruction for Z3 in Isabelle/HOL. [Електронний ресурс] – Режим доступу: <https://www21.in.tum.de/~boehmes/proofrec.pdf> (дата звернення: 29.04.2025).

Бусигіна Вероніка Павлівна – студентка групи ІПКТ-24б, кафедра комп'ютерних наук, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, м.Вінниця, e-mail: veronikabysygina@gmail.com

Науковий керівник: **Хом'юк Ірина Володимирівна** – д.пед.н., професор, професор кафедри вищої математики, Вінницький національний технічний університет, м. Вінниця, Хмельницьке шосе, 95, e-mail: vikiraivh@gmail.com

Busyhina Veronika Pavlivna – student of 1PKT-24b group, Department of Computer Science, Faculty of Intelligent Information Technology and Automation, Vinnytsia National Technical University, Vinnytsia, e-mail: veronikabysygina@gmail.com

Supervisor: **Irina V. Khomyuk** – Doctor of Science (Ped.), Professor of Higher Mathematics Department, Vinnytsia National Technical University, Vinnytsia, Khmelnytske shose, 95, e-mail: vikiraivh@gmail.com