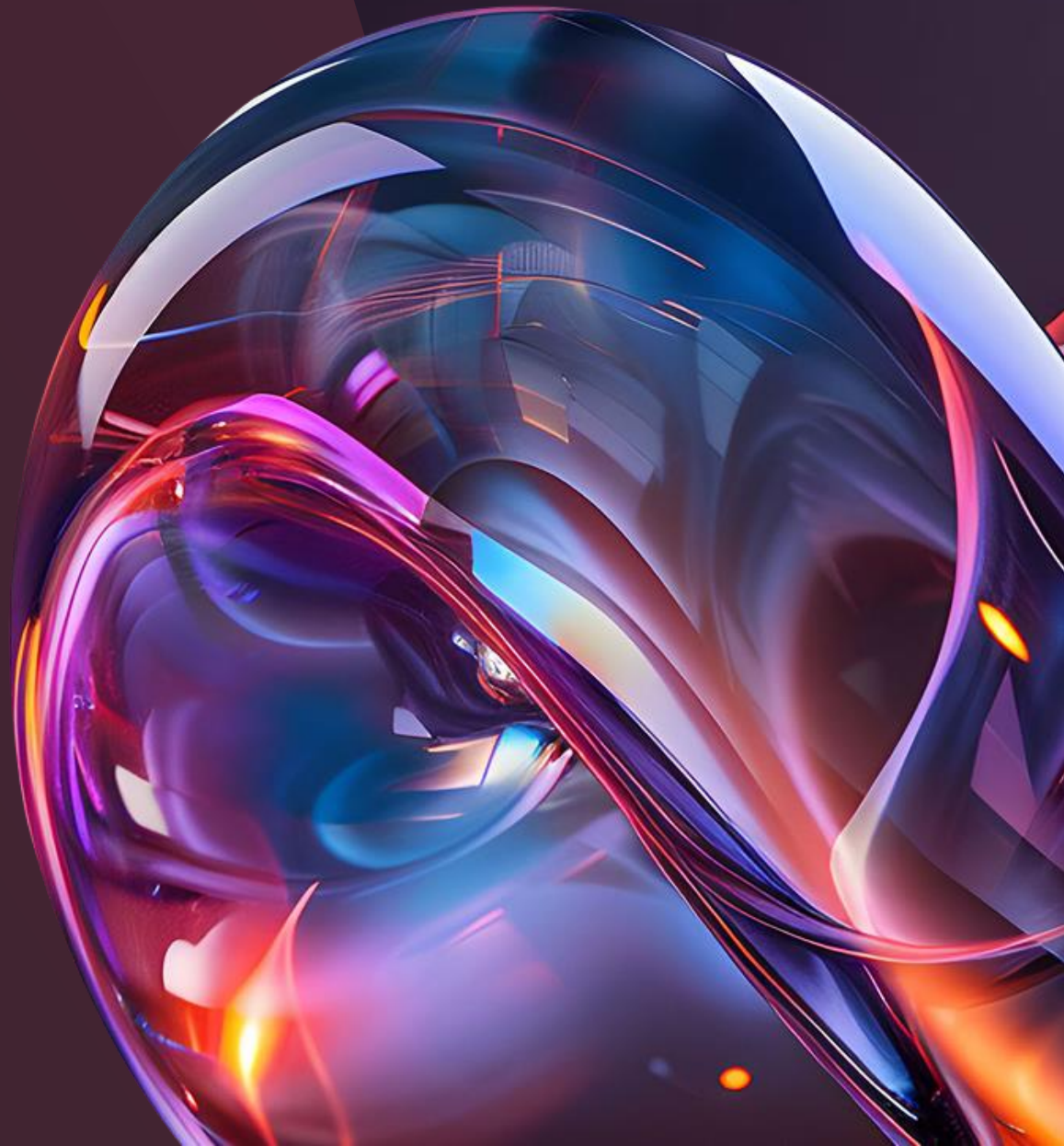
What's ahead?

Dr. Oliver Bastert
CTO, Gurobi

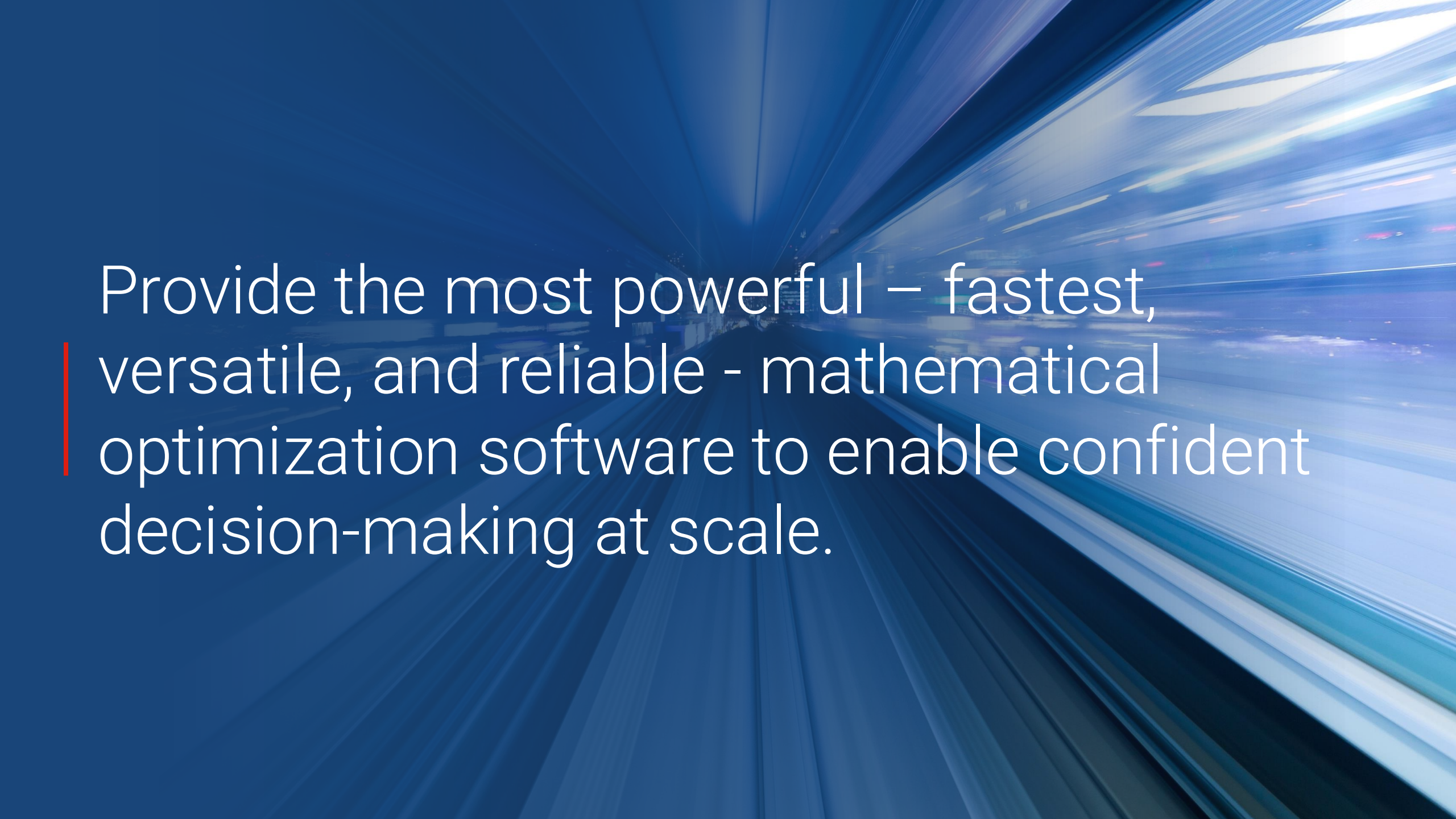


GUROBI SUMMIT

Disclaimer

This presentation includes forward-looking statements, including, but not limited to, product roadmaps, features under consideration, and anticipated timelines. These statements are for informational purposes only and do not represent a commitment, promise, or legal obligation for Gurobi Optimization, LLC (“Gurobi”) to deliver any product, feature, or functionality to a customer or prospective customer (“Customer”). The development, release, and timing of any feature or functionality described remain at Gurobi’s sole discretion and may be changed without notice.

GUROBI DOES NOT MAKE ANY PROMISES, REPRESENTATIONS, WARRANTIES OR GUARANTEES UNDER OR ARISING FROM THIS PRESENTATION. CUSTOMER SHALL NOT RELY ON ANY INFORMATION SET FORTH IN THIS PRESENTATION. CUSTOMER SHALL NOT BE ENTITLED TO ANY RECOURSE, IN THE FORM OF DAMAGES OR OTHERWISE, ARISING FROM OR RELATING TO THIS PRESENTATION.


Provide the most powerful – fastest, versatile, and reliable - mathematical optimization software to enable confident decision-making at scale.

Where does Gurobi stand out? *

GUROBI SUMMIT

01
Performance

Performance is dominating the decision. Related numerical robustness comes in fifth place.

02
Gurobi Experts

True experts working to resolve customer challenges

04
APIs

Thanks to our excellent Python API

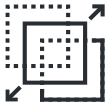
* MO Survey, 2025, Reasons for Switching to Gurobi

Agenda

For today and going forward

Performance

Performance is the #1 R&D objective.



Versatility

Expanding on the range of problem types to address more customer problems.

Acceleration

Make use of new hardware options, develop for GPUs, observe Quantum progress

Velocity

Improve algorithms to find better solutions earlier.

Integration

Expanding to making integrations and use in complex application easy.

Guidance

Utilize GenAI for developer efficiency and business user interactions.

Talent and Compute

Performance

>30%

Growth of R&D last 12 months

The most critical ingredient for creating mathematical optimization engines is excellent research & development talent.

6x

Compute capacity

Second most important is brain-power is compute capacity and our growing benchmark dataset.

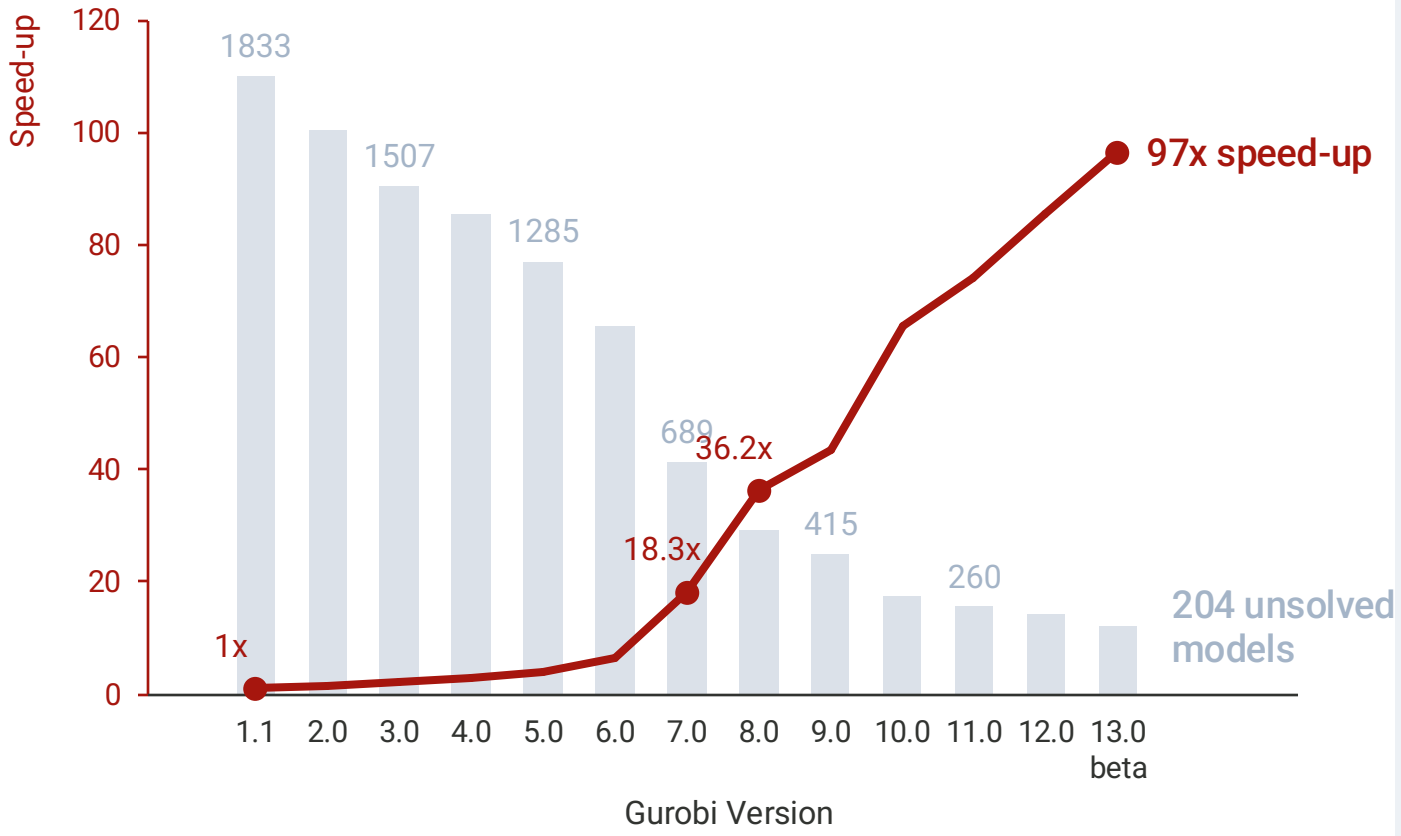
GEMS

Gurobi Engineering MIP School

We are now training our own MIP researchers & developers.

Gurobi Version Comparison: Speed and Solvability (PAR-10)

Gurobi MILP Benchmark Suite



MILP

Performance Evolution

In Gurobi's MILP benchmark suite, the latest version delivers:

- A 97x speed-up over version 1.1 in geometric mean (PAR-10) of runtimes
- Only 204 models remain unsolved after 10,000 seconds with the latest version
- The test set consists of all models that can be solved by at least one version

Time limit: 10000 sec.
Intel Xeon CPU E3-1240 v5 @ 3.50GHz
4 cores, 8 hyper-threads
32 GB RAM

Test set has 8266 models:
- 798 discarded due to inconsistent answers
- 2237 discarded that none of the versions can solve
- speed-up measured on >100s bracket: 3113 models

Nonlinear

Versatility / Performance

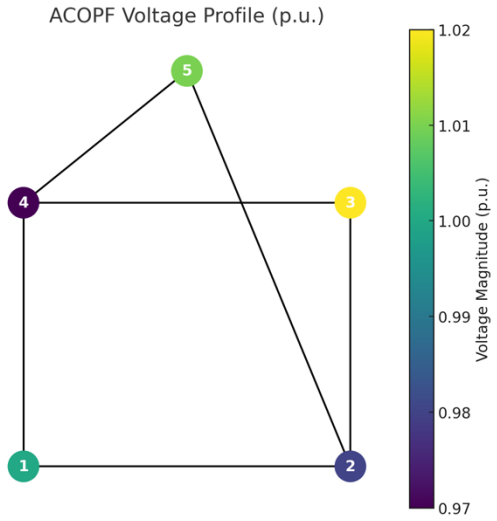
- New NL barrier implementation to
  - Solve (continuous) NLPs
  - Improve MINLP performance

- We will release the NL barrier as a preview feature in v13
- Expect NL barrier to be competitive across the board in 2026

** MINLP compared to v12*

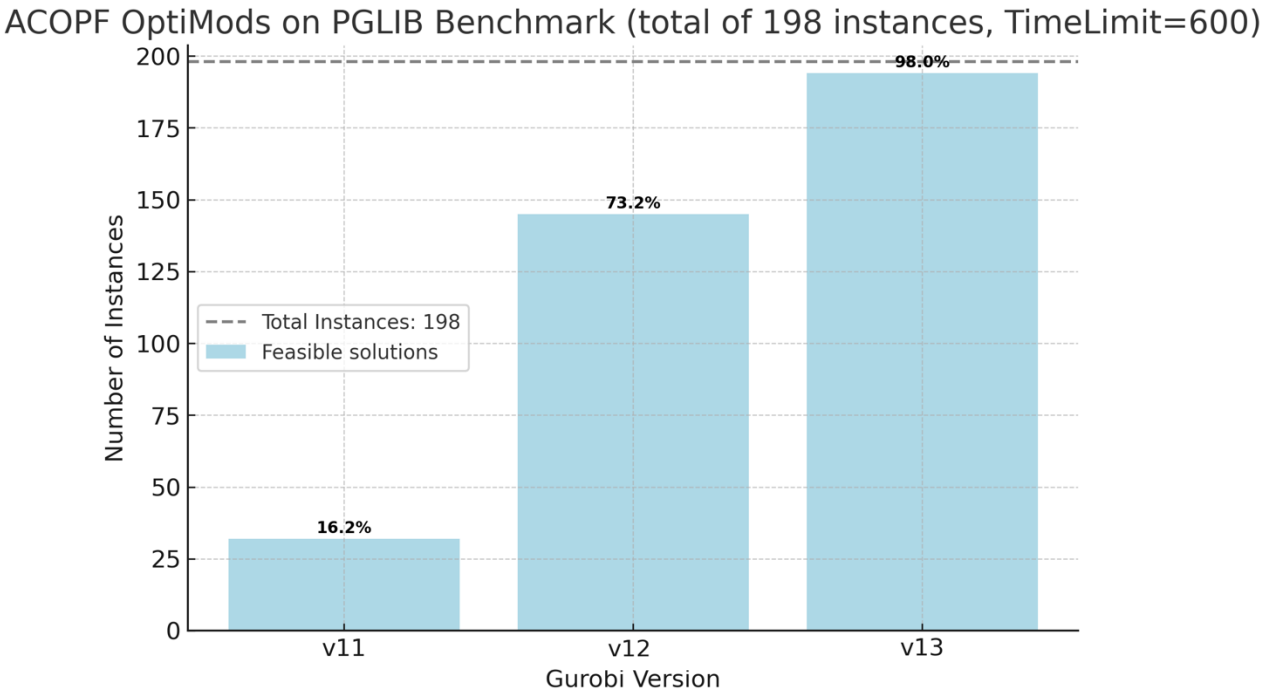
Optimal Power Flow: ACOPF (QCQP)

Versatility / Performance



Why ACOPF Matters:

- Ensures economical, secure operation of power systems
- Represents the true physics of grid operations



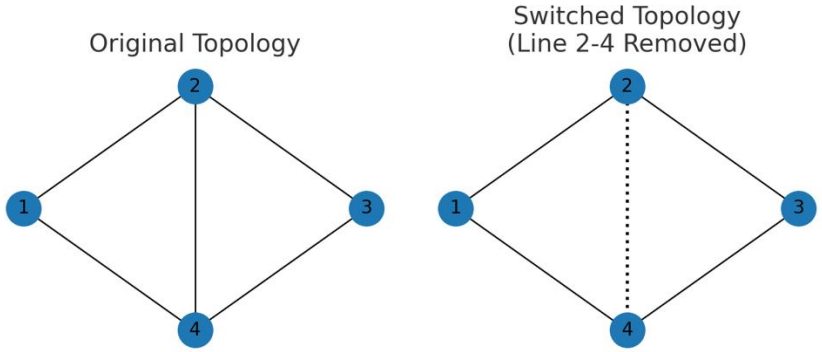
ACR: rectangular formulation (QCQP)

PGLIB Benchmarks: https://github.com/power-grid-lib/pglib-opf

Further performance enhancements expected with Polar ACOPF formulation

Optimal Power Flow: DCOPF Topology Optimization

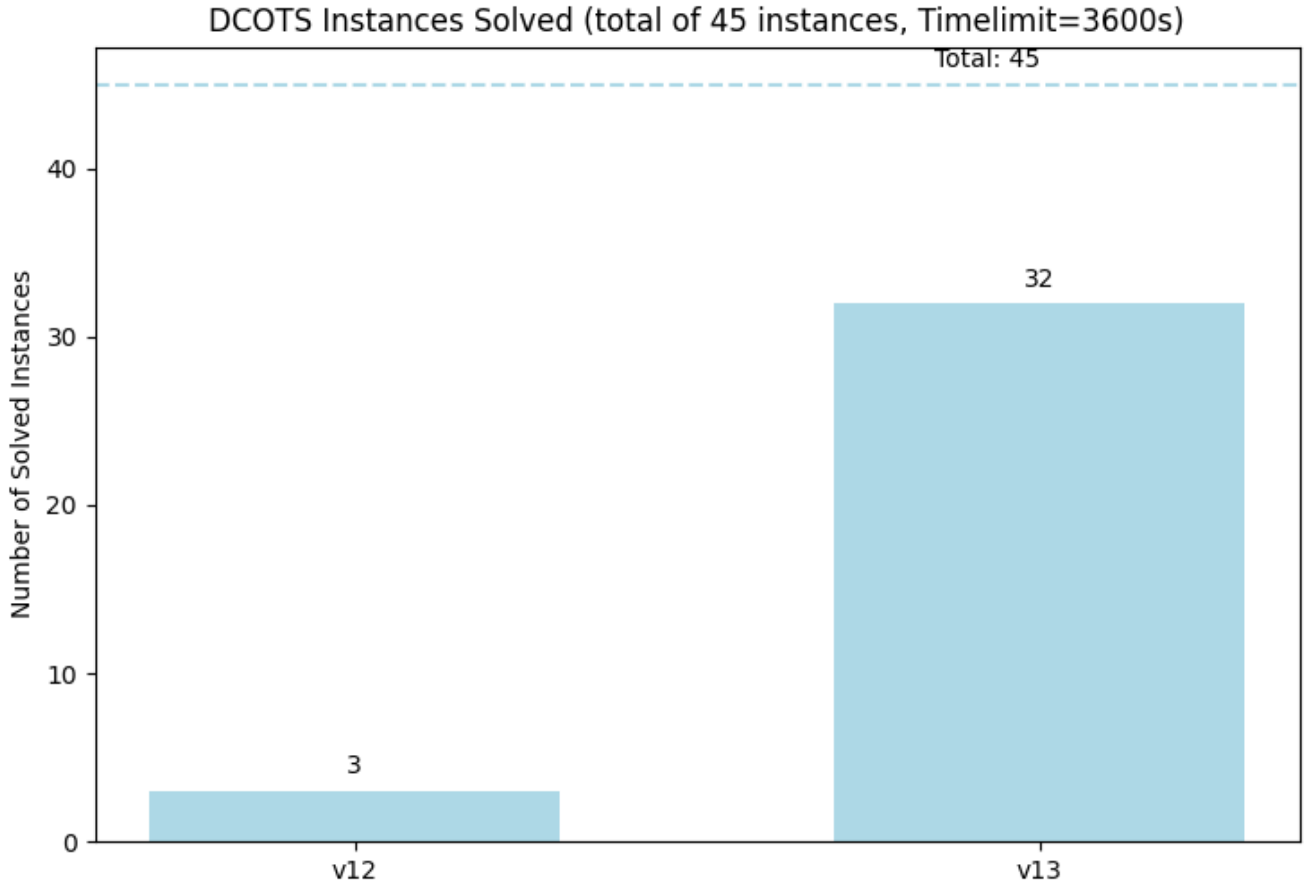
Performance



$$0 \leq p_{ij} \leq M z_{ij}$$

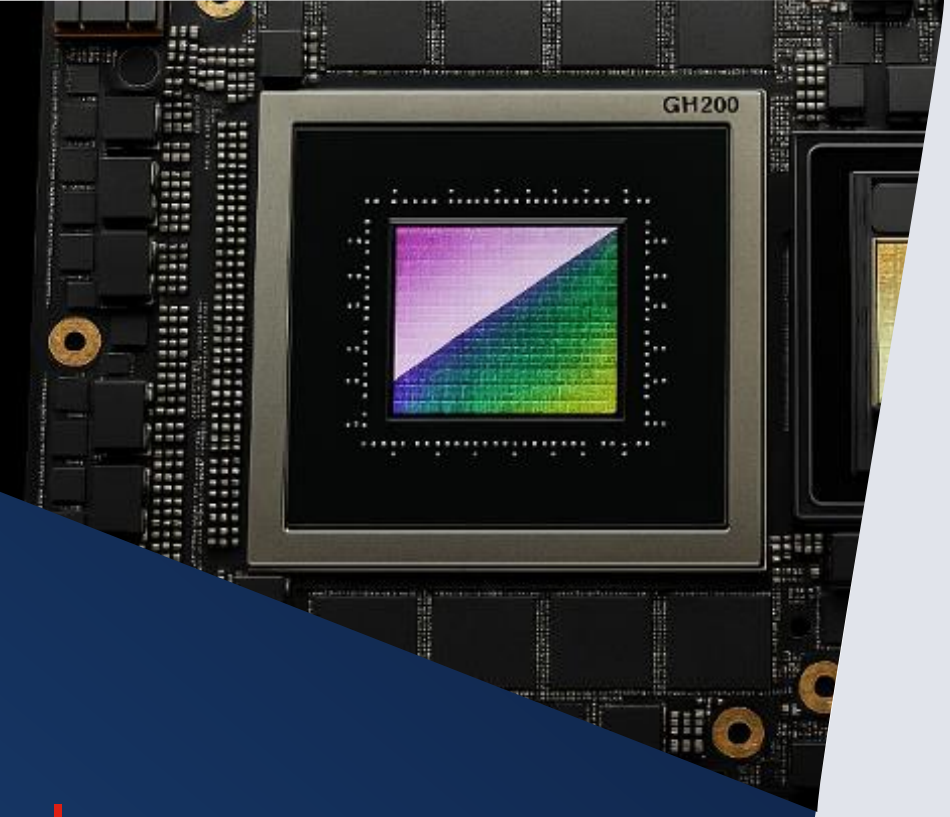
Why Topology Optimization Matters:

- Failure Recovery
- Reduce stress on congested lines (Germany alone spent EUR 4 billion on Congestion Management in 2022*)



Further benchmark results at https://openenergybenchmark.org/

* https://www.iea.org/commentaries/grid-congestion-is-posing-challenges-for-energy-security-and-transitions



LP on GPUs

Acceleration

GPU friendly algorithms required

Most existing algorithms don't work efficiently on GPUs.

Promising areas are

- Cholesky factorizations in **Barrier**
  - Limited speedup potential
- **PDHG** for solving LPs with limited accuracy
  - Basic/precise solutions require crossover
  - Increasing success on very large LPs
  - Roughly a factor of 3 difference between PDHG on CPU vs GPU*
- Heuristics & Presolve operations

Beta version of PDHG on NVIDIA hardware is available.

Both, GPU (as preview) and CPU version of PDHG are in v13, development will continue, and further opportunities will be explored.

* NVIDIA GH 200 Grace Hopper (GPU and CPU results)

GUROBI SUMMIT

Good Solutions, Quickly

Velocity

Reasons are

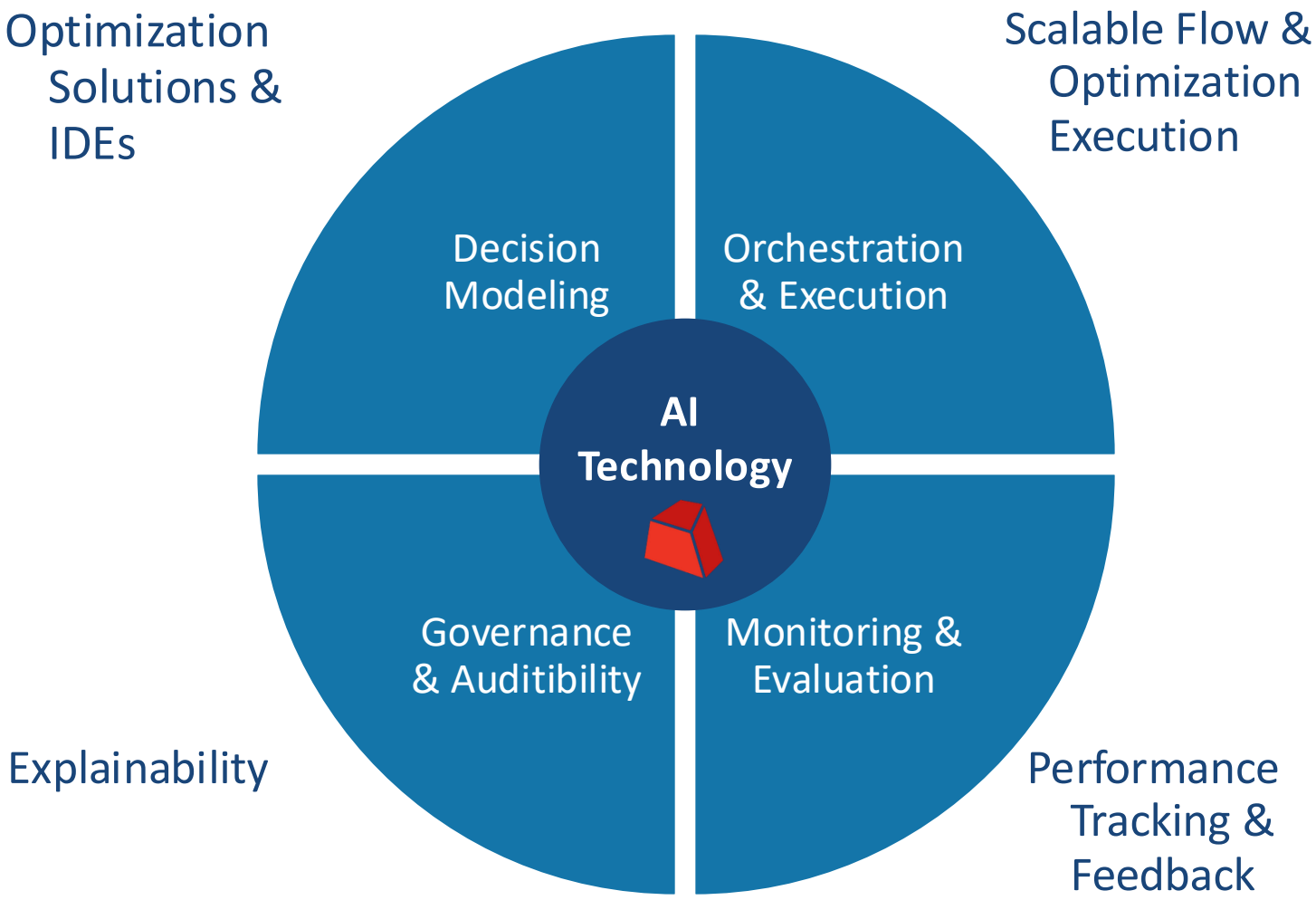
- Improving general performance
- Good solutions are good enough
- Proving optimality is hard
- Need for an (unreasonably) fast solutions

An enhanced RINS heuristic with inspiration from the Feasibility Pump will be available in v13.

Finding better solution earlier is high priority going forward.

Decision Intelligence

Integration



We don’t build a platform but …

Integration

GUROBI SUMMIT

01
Integrate with Modeling tools

Integration with classic and GenAI powered development environments

02
Expand Execution functionality

We are enabling deployment in various environments, we plan to provide more native deployment options

03
Facilitate Validation and Explainability

We will utilize GenAI for explainability, and we are exploring ideas for performance tracking and validation.

GUROBI SUMMIT

Gurobot

Guidance

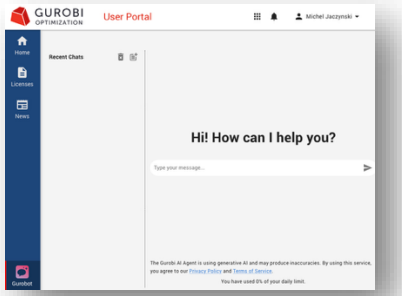
Delivers instant, expert-level assistance for its users and marks our first step in making optimization more accessible, interactive, and intelligent through Generative AI.

Gurobi Intelligence – Vision

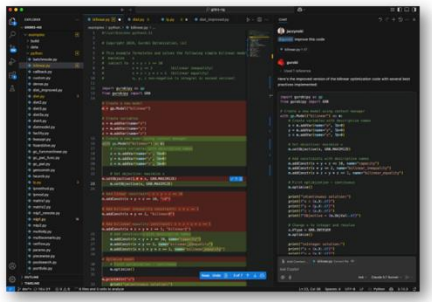
Guidance

GUROBI SUMMIT

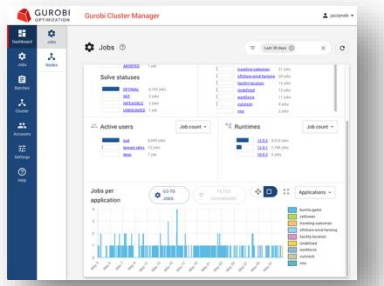
Interfaces and Integration



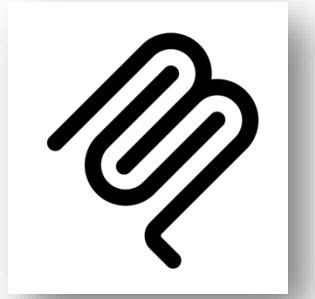
Chatbot



IDE: VSCode, Jetbrain…



Assistants in context: Cluster Manager, CLI



Interoperability: MCP, A2A

Orchestration and APIs

Gurobot

Agents for Gurobi related tasks:
Support, Tuning, Insight, Modeling, Solving…

Components

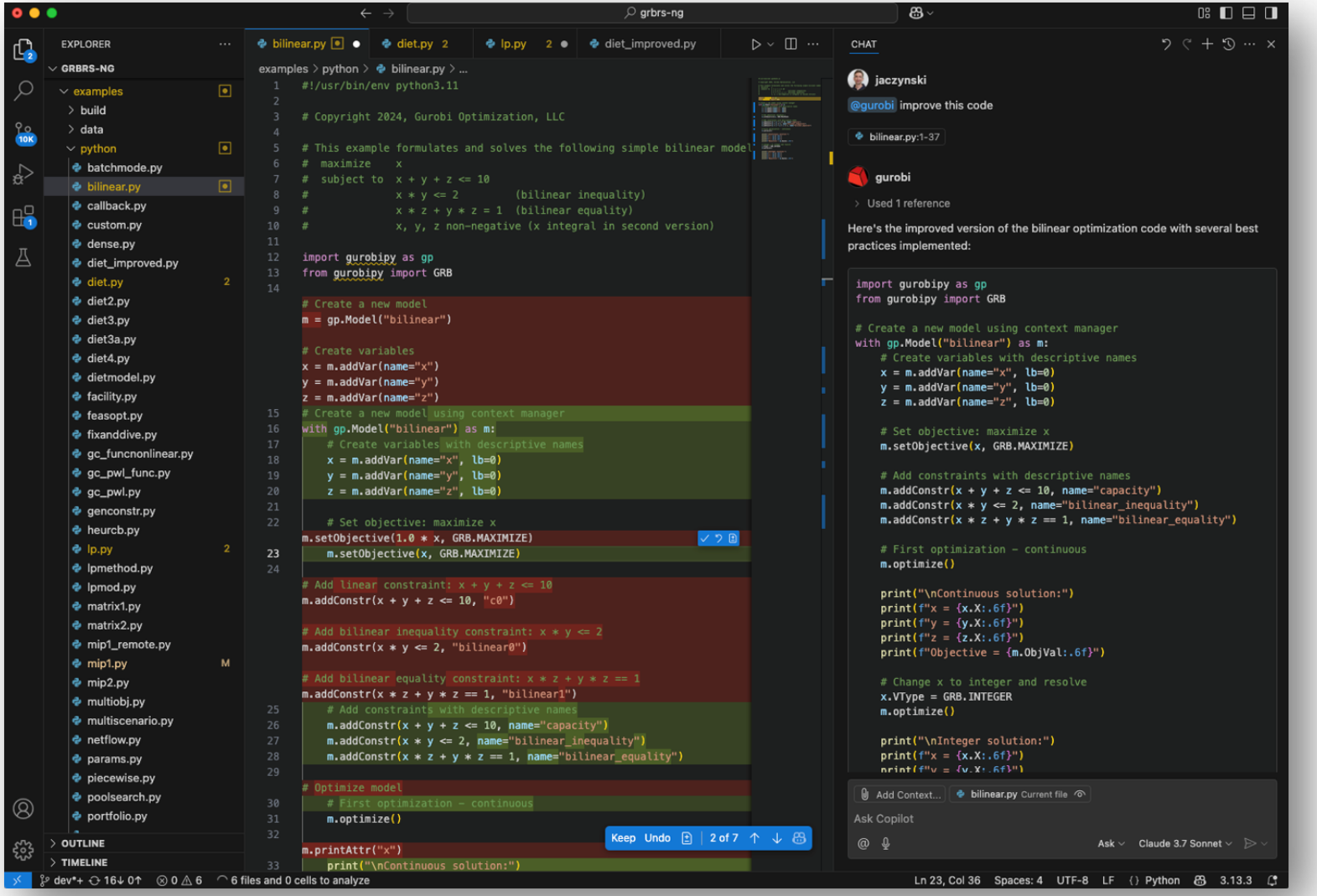
Knowledge Base to capture Gurobi best practices

Gurobi Tools:
Optimizer, Tuner, Explainer
Instant Cloud, Compute Server

LLMs:
Claude, Cohere…

Using Gurobi Intelligence in VS Code

Guidance



Spec & Test based Modeling
Guidance

GUROBI SUMMIT

User

- Give initial problem description
- Provide, review usage scenarios
- Clarify inconsistencies
- Propose, review, accept changes
- Provide and confirm updates as decision requirements change

AI

Informed Clarifications
- Suggest constraints / objectives common in similar problems
- Highlight inconsistencies, or potential omissions

Problem Description
- Assign students to classes based on preferences…

Mathematical Formulation

$$\max \sum c.x$$
$$s.t.\, A.x \leq b$$

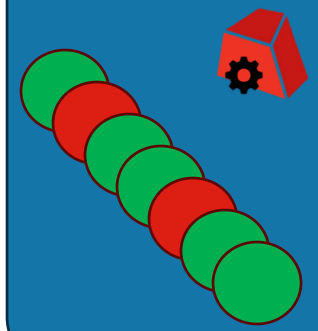
Code Implementation

```
import gurobipy as gp
m = gp.model()
```

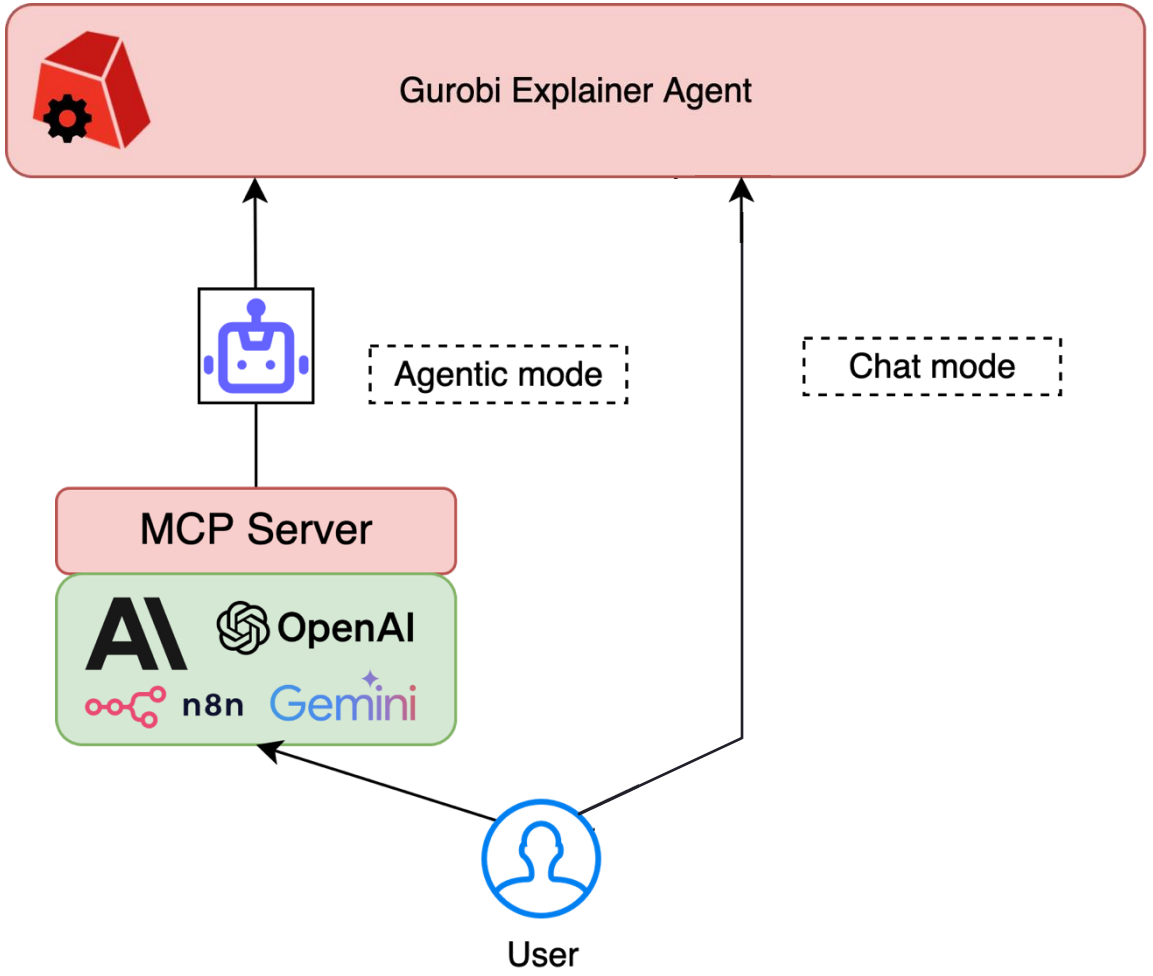
Usage Scenarios
- For a given problem, a given solution is feasible
- For a given problem, a given solution is infeasible
- For a given problem, a given solution is uniquely optimal
- For a given problem, certain characteristics must hold

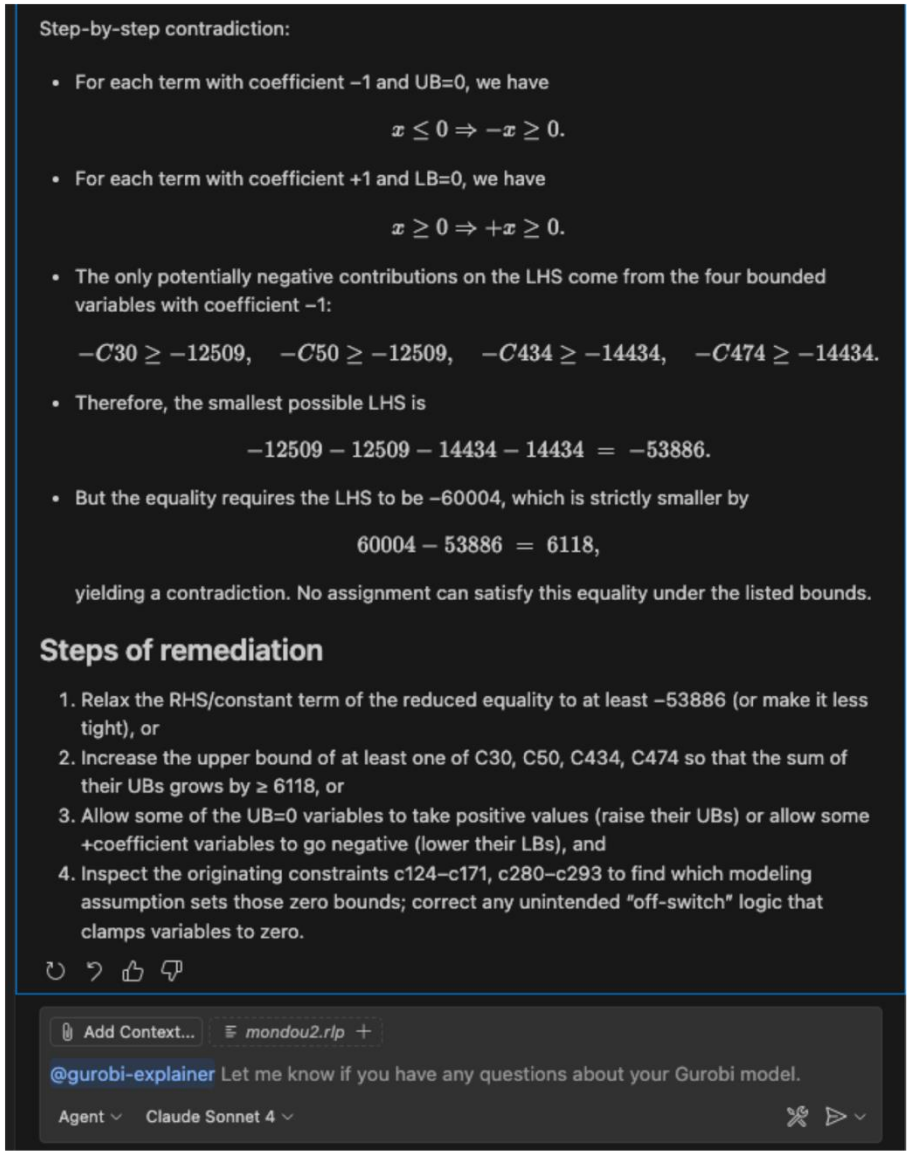
Model Correctness Tests

Results



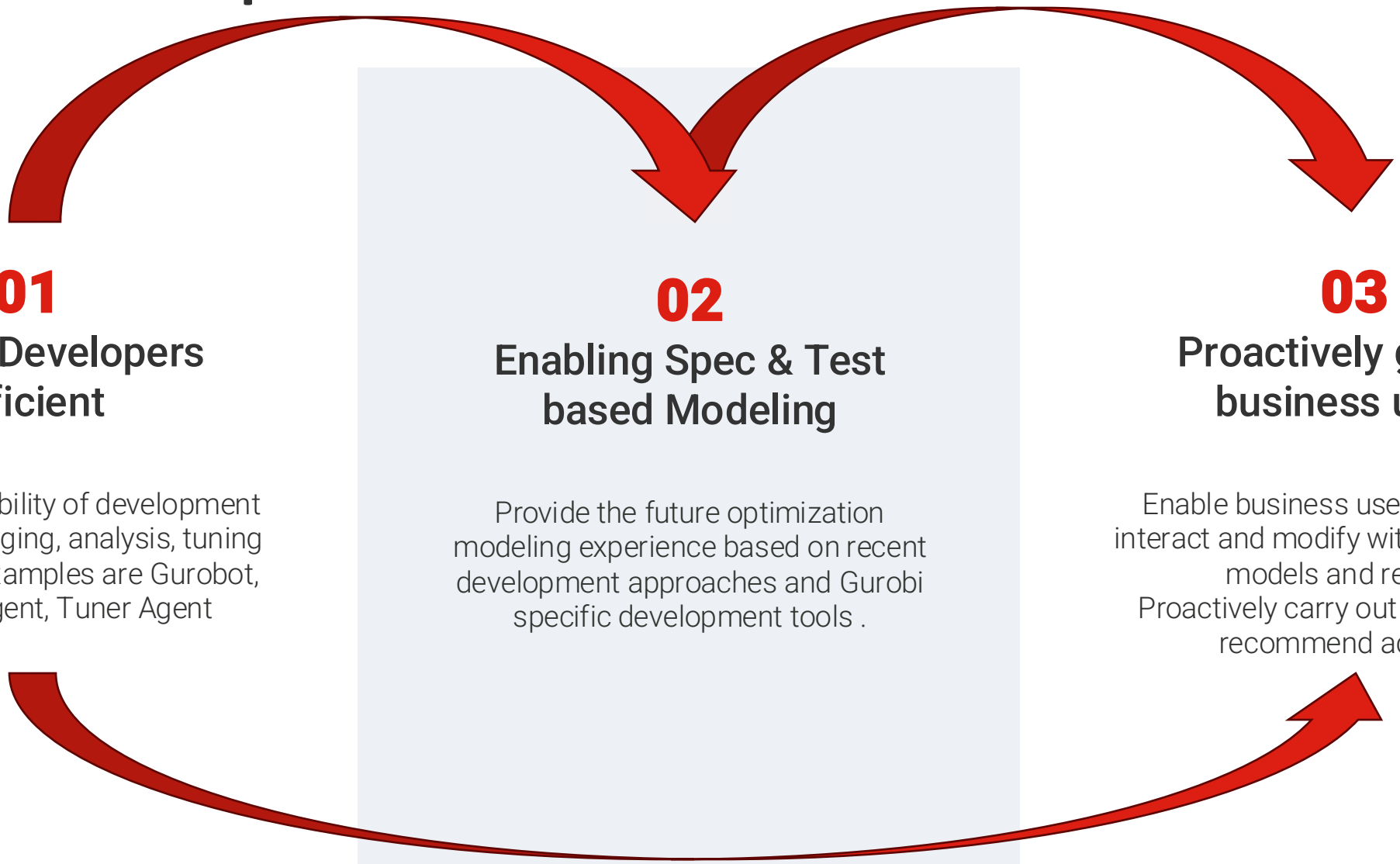
Explainer
Guidance





GenAI Roadmap

Guidance



01

Making Developers Efficient

Improve on usability of development tools, like debugging, analysis, tuning functionality. Examples are Gurobot, Explainer Agent, Tuner Agent

02

Enabling Spec & Test based Modeling

Provide the future optimization modeling experience based on recent development approaches and Gurobi specific development tools .

03

Proactively guiding business users

Enable business users to directly interact and modify with optimization models and results.
Proactively carry out analysis and recommend actions.

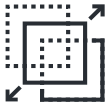
Innovation Timeline

June	July	August	September	October	November	December / January 2026
				Gurobi v13 Public Beta	Gurobi v13 Release	Gurobi 13 GPU Release
			GPU Alpha			
Gurobot Re-launch		Gurobot as First Line of Support				Explainer / MCP Available
Innovation Lab Launch	GEMS Launch					

Roadmap Summary

Performance

Performance is the #1 R&D objective.



Versatility

Expanding on the range of problem types to address more customer problems.

Acceleration

Make use of new hardware options, develop for GPUs, observe Quantum progress

Velocity

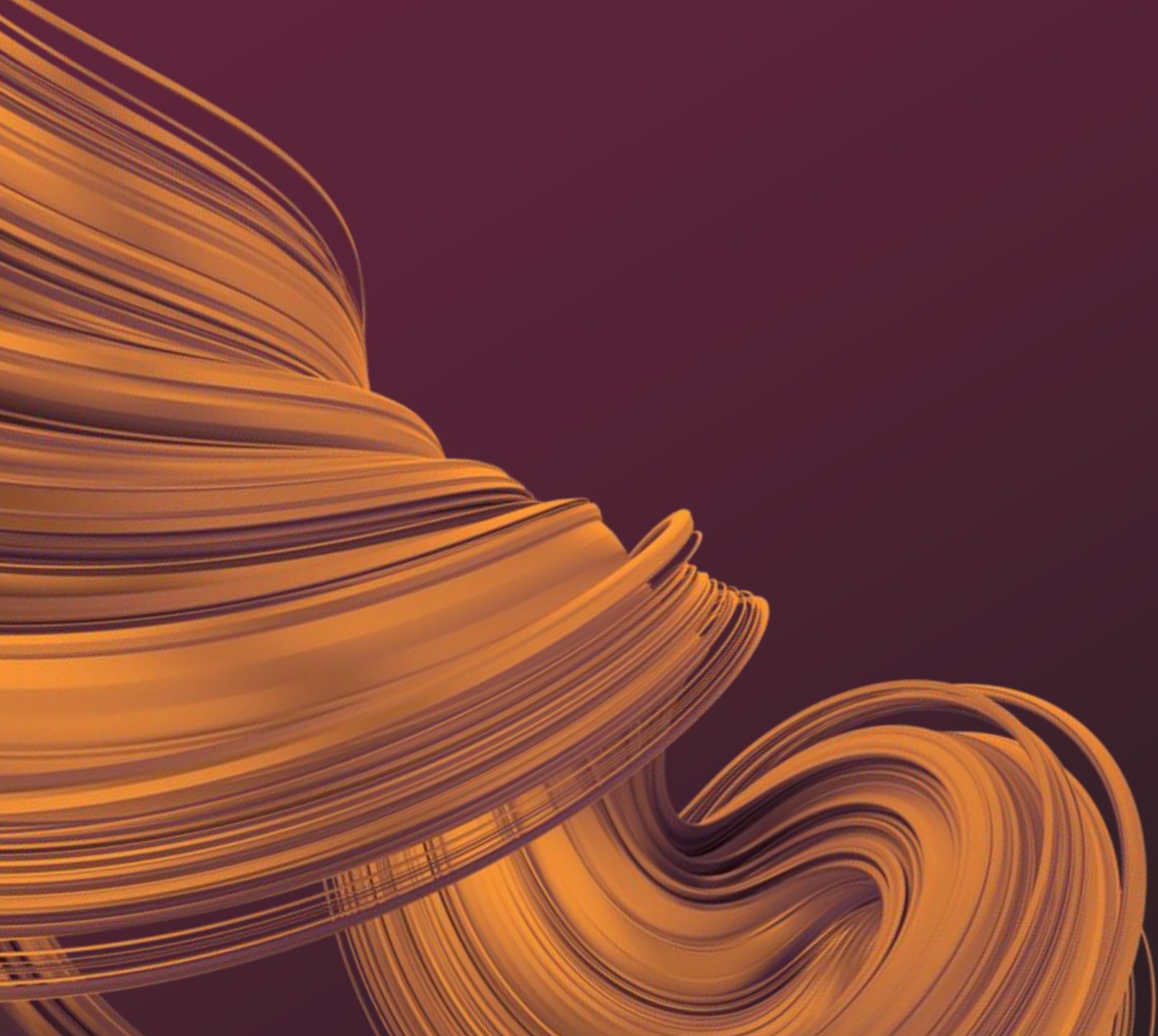
Improve algorithms to find better solutions earlier.

Integration

Expanding to making integrations and use in complex application easy.

Guidance

Utilize GenAI for developer efficiency and business user interactions.



Thank You