

BARRIERBENCH: Evaluating Large Language Models for Safety Verification in Dynamical Systems

Ali Taheri

Alireza Taban

Isfahan University of Technology, Iran

TAHERI.A@EC.IUT.AC.IR

TABAN.A@EC.IUT.AC.IR

Sadegh Soudjani

Max Planck Institute for Software Systems, Germany

University of Birmingham, UK

SADEGH@MPI-SWS.ORG

Ashutosh Trivedi

University of Colorado Boulder, USA

ASHUTOSH.TRIVEDI@COLORADO.EDU

Editors: G. Sukhatme, L. Lindemann, S. Tu, A. Wierman, N. Atanasov

Abstract

Safety verification of dynamical systems via barrier certificates is central to ensuring correctness in autonomous and other safety-critical applications. Synthesizing such certificates requires discovering mathematical functions that characterize inductive state invariants and provably separate safe and unsafe regions. Existing approaches, however, often struggle to scale computationally, depend on carefully designed templates, and rely on exhaustive or incremental searches over function spaces. They also demand substantial manual ingenuity and mathematical sophistication in constructing the search infrastructure, including selecting template families, choosing appropriate solvers, tuning hyperparameters in data-driven methods, and designing effective sampling procedures.

As a result, successful barrier certificate synthesis requires both a deep understanding of dynamical systems and control theory and practical experience with existing synthesis techniques. Much of this expertise has traditionally been transmitted among practitioners through natural language rather than formalized mathematical procedures.

This observation raises a natural question: can the linguistic and analogical reasoning that experts use informally be captured and operationalized by large language models (LLMs)? Motivated by this question, we present an *LLM-agentic framework for barrier certificate synthesis* that uses natural language reasoning to propose, refine, and validate candidate certificates. The framework combines *LLM-driven template discovery* with *SMT-based verification* and supports *barrier-controller co-synthesis* for controlled systems, thereby ensuring mathematical compatibility between safety certificates and feedback control laws.

To evaluate this capability, we introduce *BarrierBench*, a benchmark of 100 dynamical systems spanning linear, nonlinear, discrete-time, and continuous-time settings, including 68 controlled systems that require barrier-controller co-synthesis. Our experiments assess not only the effectiveness of LLM-guided barrier synthesis but also the value of *retrieval-augmented generation (RAG)* and *agentic coordination strategies* in improving reliability and performance. Across these tasks, the framework achieves over 90% success in generating valid certificates and demonstrates structural diversity, ranging from simple quadratic forms to high-order coupled polynomials. By releasing BARRIERBENCH with the accompanying toolchain, we aim to establish a *community testbed* for advancing the integration of language-based reasoning with formal verification for dynamical systems. The benchmark is publicly available at <https://hycodev.com/dataset/barrierbench>.

Keywords: Barrier Certificates, LLMs, Safety Verification, Benchmarking and Reproducibility

1. Introduction

This century has witnessed an increasingly sophisticated integration of digital software with physical infrastructure, interconnected through wired and wireless networks. These cyber-physical systems (CPS) operate so seamlessly and intuitively that it is easy to overlook the complexity beneath their surface: their successful operation depends on intricate control logic whose correctness requires a deep understanding of both the underlying physics and the discrete switching logic governing their behavior. As technological practice continues to deliver critical infrastructures such as autonomous vehicles, implantable medical devices, and smart grids, ensuring their safety has become an increasingly demanding challenge for control engineers amid ever-growing system complexity. In this paper, we address this challenge by exploring how large language models (LLMs) can assist in the synthesis and verification of safety guarantees for complex CPS.

Safety verification of dynamical systems is a canonical problem in the design of safety-critical CPS. Since safety verification is computationally undecidable even for CPS with simple piecewise-constant dynamics (Asarin and Maler, 1998), fully automatic procedures for obtaining provable safety guarantees remain out of reach. Consequently, a variety of methods has been developed to construct safety proofs, including symbolic state-space exploration, statistical guarantees, and abstraction-based techniques. Among these, the synthesis of *barrier certificates* (Prajna and Jadbabaie, 2004; Ames et al., 2019; Jagtap et al., 2021) has emerged as a particularly effective and scalable approach. Barrier certificates serve as functional analogs of inductive state invariants, characterizing the separation between safe and unsafe regions of the state space. Originating in Lyapunov’s classical theory of stability (Lyapunov, 1966; Khalil and Grizzle, 2002), they extend the central idea of certifying system behavior from convergence to invariance. Their formulation as the search for a single scalar function satisfying suitable inequality constraints makes them amenable to optimization- and satisfiability-based methods, contributing to their prominence in both theory and practice. We therefore focus on barrier-certificate-based proof discovery as a promising foundation for integrating formal reasoning with language-model-guided synthesis.

A **barrier certificate** is a scalar function whose level sets define an inductive invariant separating unsafe states from all trajectories originating from admissible initial conditions. By providing formal guarantees over infinite time horizons, barrier certificates are central to correctness arguments for both linear and nonlinear dynamical systems. Traditional approaches to barrier synthesis are predominantly optimization-based. Sum-of-squares and semidefinite programming methods cast the search as a convex optimization problem over polynomial templates (Prajna et al., 2007; Kordabad et al., 2025). These methods provide strong correctness guarantees, but their computational cost grows rapidly with system dimension, and they are typically limited to low-degree polynomial classes. To improve generality, counterexample-guided inductive synthesis (CEGIS) frameworks—see Dawson et al. (2023) for a comprehensive review—combine such methods with Satisfiability Modulo Theories (SMT) solvers to iteratively refine candidate barriers (Ravanbakhsh and Sankaranarayanan, 2015). Neural and data-driven variants use deep architectures to approximate more expressive certificate functions (Abate et al., 2024; Neustroev et al., 2025; Abate et al., 2021). Although these approaches broaden the space of representable barriers, they also introduce new challenges, including training instability, reliance on finite samples, reduced interpretability, and continued dependence on carefully chosen templates or architectures. As a result, successful synthesis still depends heavily on expert intuition in selecting suitable function classes, solvers, and hyperparameters. These decisions often reflect informal reasoning that remains difficult to encode within existing algorithmic frameworks.

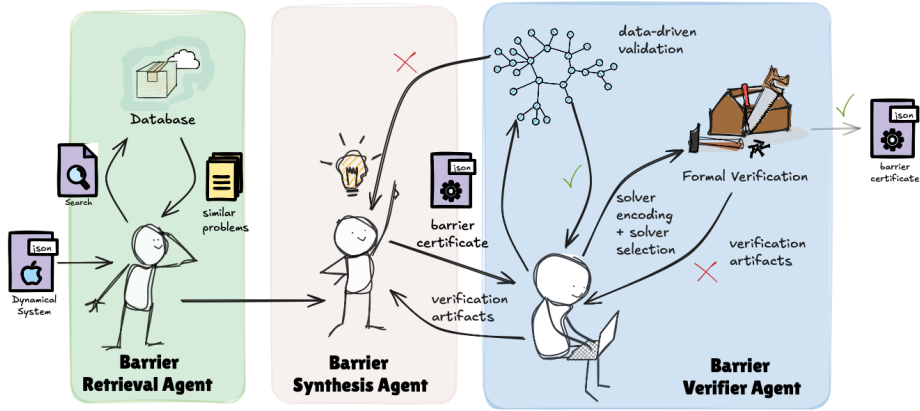


Figure 1: Agentic framework for LLM-guided barrier certificate synthesis.

This paper investigates whether LLMs can capture and operationalize the linguistic and analogical reasoning that practitioners use when constructing barrier certificates.

We present an LLM-guided synthesis framework for barrier certificates that leverages natural-language reasoning to propose, refine, and validate candidate certificates. The framework combines *LLM-driven template discovery* with *SMT-based formal verification* and naturally extends to *barrier-controller co-synthesis* for controlled systems. By incorporating verification feedback into subsequent LLM prompts, the framework supports dynamic exploration of barrier templates and coefficients, often uncovering novel mathematical forms that fixed-template approaches (Prajna et al., 2007; Kong et al., 2013; Ames et al., 2019; Abate et al., 2021; Kordabad et al., 2025) may overlook. LLMs guide the synthesis process by reasoning about system properties, safety requirements, and mathematical structure, first analyzing local safety conditions and then integrating them into global certificates. This integration provides the expressiveness and adaptability needed for complex systems while preserving the mathematical rigor required in safety-critical applications.

To operationalize this idea, we design an *agentic architecture* that emulates the workflow of expert control designers. Human designers typically draw on prior experience with similar dynamical systems to select barrier templates, solver strategies, and proof heuristics. Analogously, our framework instantiates three collaborating agents: a *Barrier Retrieval Agent*, which searches a database of previously solved systems to identify analogous examples; a *Barrier Synthesis Agent*, which proposes candidate barrier certificates inspired by retrieved analogs; and a *Barrier Verifier Agent*, which evaluates these candidates in two stages, first through a lightweight data-driven check and then through formal SMT-based verification. Feedback from the Verifier Agent is returned to the Synthesis Agent, forming a closed-loop interaction bounded by a global timeout. This design enables iterative refinement of candidate certificates while making effective use of external formal tools. Our study investigates the following research questions:

- RQ1:** Are state-of-the-art large language models already capable of independently designing valid barrier certificates for safety verification tasks?
- RQ2:** Does knowledge of barrier certificates for similar systems improve synthesis performance, and can retrieval-augmented generation (RAG) facilitate this transfer?
- RQ3:** Can integrating specialized agents—retrieval, synthesis, and verification—into an agentic framework that interfaces with external solvers improve the efficiency and reliability of LLM-guided barrier synthesis?

To evaluate these questions, we introduce *BarrierBench*, a benchmark of 100 dynamical systems spanning linear, nonlinear, discrete-time, and continuous-time settings, including 68 controlled systems that require barrier-controller co-synthesis. Across these tasks, the LLM framework achieves over 90% success in generating valid certificates and exhibits structural diversity ranging from simple quadratics to high-order coupled polynomials. By releasing BARRIERBENCH and the accompanying toolchain, we aim to establish a community testbed for advancing the integration of language-based reasoning with formal verification for dynamical systems.

2. Barrier Certificates

A dynamical system is defined as a tuple $\mathcal{S} = (X, f, X_I, X_U)$, where $X \subseteq \mathbb{R}^n$ is the state space, $f : X \rightarrow \mathbb{R}^n$ is a vector field, $X_I \subseteq X$ is the set of initial states, and $X_U \subseteq X$ is the set of unsafe states with $X_I \cap X_U = \emptyset$. For continuous-time systems, the evolution is governed by

$$\dot{x}(t) = f(x(t)), \quad x(0) \in X_I, \quad (1)$$

where $x(t) \in X$ denotes the state at time $t \geq 0$. For discrete-time systems, the evolution is given by

$$x[k+1] = f(x[k]), \quad x[0] \in X_I, \quad (2)$$

where $k \in \mathbb{N}$ is the discrete time step.

Definition 1 (Safety Specification) *Given a dynamical system $\mathcal{S} = (X, f, X_I, X_U)$, \mathcal{S} is safe if trajectories starting in X_I remain in $X \setminus X_U$ for all time.*

Definition 2 (Barrier Certificate) *A barrier certificate is a scalar function $B : X \rightarrow \mathbb{R}$ satisfying:*

$$B(x) \leq 0, \quad \forall x \in X_I, \quad (3)$$

$$B(x) > 0, \quad \forall x \in X_U, \quad (4)$$

$$\nabla B(x) \cdot f(x) < 0, \quad \forall x \in X \text{ s.t. } B(x) = 0 \text{ (continuous-time),}$$

$$\text{or } B(f(x)) - B(x) \leq 0, \quad \forall x \in X \text{ (discrete-time).} \quad (5)$$

Here, $\nabla B(x) \cdot f(x)$ denotes the Lie derivative $\mathcal{L}_f B(x)$ of a continuously differentiable function $B(x)$ along the vector field $f(x)$, defined as $\mathcal{L}_f B(x) = \sum_{i=1}^n \frac{\partial B}{\partial x_i} f_i(x)$.

Theorem 3 (Safety via Barrier Certificates (Prajna and Jadbabaie, 2004)) *If there exists a barrier certificate $B(x)$ for \mathcal{S} , then the system \mathcal{S} is safe.*

The above definitions and theorem are stated for systems without control inputs. If the system has control input $u \in U$, the objective is to design a state feedback law that gives u as a function of x and makes the system safe. For this, the barrier conditions are quantified existentially over the input.

3. Agentic Framework for Barrier Synthesis

Our synthesis framework employs a multi-agent architecture in which specialized LLM-powered agents collaborate to discover barrier certificates. The framework comprises three core agents operating in an iterative pipeline: (1) a *Barrier Retrieval Agent* that identifies relevant prior solutions; (2) a *Barrier Synthesis Agent* that generates candidate barrier certificates; and (3) a *Barrier Verifier Agent* that checks candidates and provides feedback for refinement. This iterative process addresses limitations of conventional approaches by enabling informed template selection and adaptive exploration of diverse mathematical structures through verification-guided synthesis.

Framework Pipeline. The framework \mathcal{F} takes as input a dynamical system (X, f, X_I, X_U) , with optional control input U for controlled systems. A persistent database stores previously verified barrier certificates, which the Retrieval Agent accesses to accelerate synthesis on similar problems. The pipeline consists of four stages: (1) historical retrieval, (2) LLM-based synthesis, (3) two-stage verification, and (4) iterative refinement guided by feedback. The system performs up to K main iterations, each exploring new template structures, and up to R refinements per iteration to adjust coefficients or terms. The process terminates upon discovery of a valid barrier certificate or exhaustion of all attempts. If no valid certificate is found, the framework returns the candidate maximizing the proportion of satisfied conditions. Figure 1 illustrates the overall architecture of \mathcal{F} .

Barrier Retrieval Agent. This agent analyzes the input system and retrieves relevant examples from the database through a two-stage process. First, it performs feature-based retrieval using automatically extracted system attributes such as dimension, time domain, linearity, and topology. Candidate problems matching all features are shortlisted. Second, an LLM-based reasoning stage ranks these candidates for structural similarity, providing the most compatible example to the Synthesis Agent while noting that adaptation may be required. When no match is found, the agent instructs the Synthesis Agent to proceed independently. As shown in Section 4.4, this retrieval step significantly accelerates convergence by enabling frequent first-try success.

Barrier Synthesis Agent. This agent leverages the reasoning abilities of LLMs to analyze system dynamics and propose barrier certificates that satisfy the safety conditions (3)–(5). During the first main iteration ($k = 1$), it incorporates relevant historical examples from the Retrieval Agent. For later iterations and refinements, it integrates feedback from the Verifier Agent to modify structure and coefficients. For systems requiring control synthesis, the agent jointly produces both the barrier certificate and controller expressions to ensure compatibility between safety and control objectives.

Barrier Verifier Agent. The Barrier Verifier Agent conducts a two-stage evaluation of candidate barriers. In the first stage, it performs a sample-based validation across N points drawn from X_I , X_U , and X , where N is selected to balance computational efficiency with adequate state space coverage; we use $N = 5000$ in our experiments to achieve this trade-off, discarding invalid barriers that violate any of the three conditions on at least one of sampled points, to avoid unnecessary SMT solver calls. In the second stage, it performs formal verification using SMT solvers such as Z3 (de Moura and Bjørner, 2008), Yices (Dutertre, 2014), and cvc5 (Barbosa et al., 2022), checking the unsatisfiability of negated barrier conditions. The agent selects solvers adaptively based on problem structure (e.g., polynomial vs. transcendental terms), expression complexity, and implicit knowledge of solver capabilities (e.g., whether transcendental functions are supported), and then applies symbolic conversions or Taylor approximations when necessary, and handles verification failures through solver switching or adaptive timeout adjustments.

Iterative Refinement Mechanism. When verification fails, the Verifier Agent provides the violated conditions and counterexamples as feedback to the Synthesis Agent, which uses them to refine the barrier. In early refinements ($r \leq 2$), the agent adjusts coefficients while preserving structure; in later refinements ($r > 2$), it performs structural modifications such as adding or removing coupling terms while maintaining polynomial degree. This strategy balances exploitation (fine-tuning promising templates) and exploration (structural variation across iterations). The Verifier Agent logs all intermediate results, enabling cumulative feedback that helps the Synthesis Agent learn from previous

Algorithm 1 Agentic Barrier Certificate Synthesis**Input:** System (X, f, X_I, X_U) ; max iterations K ; refinements R ; input set U for controlled systems**Output:** Barrier B^* , controller C^* for controlled systems, or failureInitialize: $\text{best_score} \leftarrow 0$; $B^*, C^* \leftarrow \text{null}$

```

for  $k \leftarrow 1$  to  $K$  do
   $e \leftarrow \text{RETRIEVALAGENT}(f, X_I, X_U, \mathcal{D})$   $(B_k, C_k) \leftarrow \text{SYNTHESISAGENT}(e, \text{feedback}_k)$ 
  if  $\text{VERIFIERAGENT.SAMPLECHECK}(B_k, C_k)$  fails then
    | continue
  end
   $f' \leftarrow f$  with  $C_k$ ;  $(\text{valid}_k, \text{feedback}_k) \leftarrow \text{VERIFIERAGENT.FORMALCHECK}(B_k, f')$ 
  if  $\text{valid}_k$  then
    |  $\text{RETRIEVALAGENT.STORE}(B_k, C_k)$  return  $(B_k, C_k)$ 
  end
  for  $r \leftarrow 1$  to  $R$  do
     $(B_k^r, C_k^r) \leftarrow \text{SYNTHESISAGENT.REFINE}(\text{feedback}_k, r)$ 
    if  $\text{VERIFIERAGENT.SAMPLECHECK}(B_k^r, C_k^r)$  fails then
      | continue
    end
     $f' \leftarrow f$  with  $C_k^r$ ;  $(\text{valid}_k^r, \text{feedback}_k^r) \leftarrow \text{VERIFIERAGENT.FORMALCHECK}(B_k^r, f')$ 
    if  $\text{valid}_k^r$  then
      |  $\text{RETRIEVALAGENT.STORE}(B_k^r, C_k^r)$  return  $(B_k^r, C_k^r)$ 
    end
  end
end
return Best partial solution

```

attempts. As shown in Section 4.4, this mechanism substantially increases success rates, with a large number of cases solved within one or two main iterations.

4. Barrierbench: Experiments and Case Studies

The complete set of prompts used in our framework are included in the Appendix A.

4.1. Experimental Setup

The framework operates with carefully chosen hyperparameters validated across various systems. We set maximum main iterations $K = 5$ and maximum refinements per iteration $R = 4$. The LLM utilizes Claude Sonnet 4 and ChatGPT-4o with engineered prompts designed to synthesize mathematical expressions while maintaining formal correctness. All experiments use Z3 version 4.15.1, cvc5 version 1.3.1, and Yices version 2.6.4 as the SMT solvers for formal verification. For the verifier agent, we use in total $N = 5000$ samples with 30% from the initial set, 30% from the unsafe set, and 40% from the state space.

4.2. Case Studies

We evaluate our agentic barrier synthesis framework on BARRIERBENCH, a benchmark comprising 100 dynamical systems. Our evaluation case studies encompasses continuous-time systems, discrete-time systems, and controlled systems requiring simultaneous barrier-controller co-synthesis, with a variety of complex nonlinear dynamics across dimensions from 1D to 8D.

BARRIERBENCH includes 68 controlled systems for which our framework demonstrates the simultaneous synthesis of barrier certificates and controllers parameters, which together establish safety guarantees for closed-loop dynamics. The characteristics of our benchmark are shown in Table 1.

Table 1: Dataset characteristics of BARRIERBENCH.

| Characteristic | Description |
|----------------------|--------------------------------------|
| Dimensions | 1D–8D |
| Initial set topology | Ball, rectangle |
| Unsafe set topology | Ball, rectangle, union of rectangles |
| System types | Controlled, autonomous |
| Time domain | Continuous-time, discrete-time |
| Dynamics types | Linear, nonlinear |

4.3. Qualitative Results

In this section, we qualitatively review the results of some of the benchmarks synthesized by our framework, for more case studies you can check <https://hycodev.com/dataset/barrierbench>.

- **Continuous-Time Systems.** The 4D system $\dot{x}_1 = -x_1 + 4.5, \dot{x}_2 = x_1 - x_2 - 3, \dot{x}_3 = -0.5x_3, \dot{x}_4 = -0.3x_4$ with state space \mathbb{R}^4 and polynomial degree of 1 has rectangular initial set $[1.75, 2.25] \times [1.75, 2.25] \times [-0.1, 0.1] \times [-0.1, 0.1]$ and rectangular unsafe region $[9, 10] \times [9, 10] \times [-5, 5] \times [-5, 5]$. The framework generates the quadratic barrier $B(x) = (x_1 - 4.5)^2 + (x_2 - 1.5)^2 + x_3^2 + x_4^2 - 25$ of degree 2.
- **Discrete-Time Systems.** The 2D system $x_1[k+1] = x_1[k] + 0.01(-100x_1[k] - x_2[k]), x_2[k+1] = x_2[k] + 0.01(x_1[k] - 100x_2[k])$ with state space \mathbb{R}^2 and polynomial degree $d_f = 1$ has rectangular initial set $[0.1, 0.4] \times [0.1, 0.55]$ and rectangular unsafe region $[0.45, 0.5] \times [0.6, 1.0]$. The framework generates the quadratic barrier $B(x) = x_1^2 + x_2^2 - 0.5$ of degree 2.
- **Controlled Systems with Barrier-Controller Co-Synthesis.** The 4D system $\dot{x}_1 = x_2 + u_0, \dot{x}_2 = -0.95 \sin(x_1) - 0.02x_2 + 0.01x_3 + u_1, \dot{x}_3 = -0.1x_3 + 0.02x_1 + u_2, \dot{x}_4 = -0.03x_4 + 0.01x_2 + u_3$ with state space \mathbb{R}^4 , control space \mathbb{R}^4 , and non-polynomial dynamics has initial ball with center $(0, 0, 0, 0)$ and radius 0.3, and unsafe set as the complement of the ball with center $(0, 0, 0, 0)$ and radius 2.5. The framework synthesizes controllers $u_0 = -3x_1 - 1.5x_2, u_1 = -3x_2 + 0.8 \sin(x_1), u_2 = -3x_3 - 0.05x_1, u_3 = -3x_4 - 0.02x_2$ with quadratic barrier $B(x) = x_1^2 + x_2^2 + x_3^2 + x_4^2 - 4.0$ of degree 2.

4.4. Ablation Study

In this section, we compare our full pipeline against a baseline that uses only a single LLM prompt without any framework support to evaluate our framework. Moreover, to evaluate the individual contributions of our framework components, we conduct ablation experiments. Table 2 presents the comparison between our framework and a single LLM prompt based on the number of solved problems in BARRIERBENCH.

The baseline approach, which relies on a single prompt LLM generation without any barrier retrieval from the historical database, iterative refinement, or failure-guided refinement, successfully solves only 41 problems (41%) with Claude Sonnet 4 and only 17 problems (17%) with ChatGPT-4o. In contrast, our full framework achieves substantial improvements on both models: 90% success rate with Claude Sonnet 4 and 46% success rate with ChatGPT-4o, representing improvements of +49 and +29 percentage points, respectively. This performance gap validates the significant role of our framework components and provides a negative answer to **RQ1**. The baseline’s limited success rates

Table 2: Comparison of baseline (single-prompt) and agentic framework on BARRIERBENCH.

| Approach | Claude Sonnet 4 | | ChatGPT-4o | |
|--------------------------|-----------------|---------------|--------------|---------------|
| | Success Rate | Solved | Success Rate | Solved |
| Baseline (Single Prompt) | 41% | 41/100 | 17% | 17/100 |
| Full Framework | 90% | 90/100 | 46% | 46/100 |
| <i>Improvement</i> | +49% | +49 | +29% | +29 |

Table 3: Number of problems solved at each iteration–refinement combination. The index of I shows the number of iterations and the index of R shows the number of refinements. For example, the number under I_1-R_0 denotes the number of cases solved in first-try without refinement, leveraging prior database search for similar problems.

| | R_0 | R_{1-2} | R_{3-4} | Total |
|--------------|------------|------------|------------|------------|
| I_1 | 49 | 9 | 5 | 63 (70.0%) |
| I_2 | 10 | 3 | 2 | 15 (16.7%) |
| I_{3+} | 4 | 2 | 6 | 12 (13.3%) |
| Total | 63 (70.0%) | 14 (15.6%) | 13 (14.4%) | 90 (100%) |

demonstrate that state-of-the-art LLMs cannot independently design valid barrier certificates without structured framework support. To further analyze the distribution of synthesized valid barriers across main iterations and refinements, Table 3 presents a detailed breakdown showing when and how problems are solved within our framework.

Barrier Retrieval Impact. As shown in Table 3, our barrier retrieval mechanism, which combines feature-based filtering and LLM selection accelerates convergence. Among the 90 successfully solved problems, 49 (54.4%) are solved in the first attempt without any refinements ($I_1 - R_0$), compared to 41 total successes from the baseline. This indicates that our intelligent barrier retrieval not only increases the overall success rate but also accelerates solution discovery. Problems that require multiple iterations or fail entirely under the baseline approach are now solved by leveraging patterns from problems with matching properties, which provides a positive answer to **RQ2**.

Our retrieval does not transfer certificates directly. It provides structural templates as few-shot examples for the LLM. Similar to few-shot learning (Brown et al., 2020), the synthesis agent uses architectural patterns from analogous problems, then generates entirely new coefficients and controllers tailored to the current problem’s specific constraints while explicitly instructed not to copy solutions (see Appendix A). For out-of-distribution scenarios where retrieval fails, the framework gracefully degrades to independent synthesis (baseline 41% capability) with iterative structural exploration. Importantly, as discussed in Section 3, failures from initial attempts accumulate as feedback, constrain the exploration space by showing what does not work, enabling the synthesis agent to propose progressively better structures in subsequent iterations.

Refinement Impact. The refinement mechanism contributes substantially to the framework performance. As shown in Table 3, while 63 problems (70%) require no refinement (R_0 column), an

additional 14 problems (15.6%) benefit from non-structural refinement (R_{1-2}), and 13 problems (14.4%) require structural adjustments (R_{3-4}). Within the first iteration alone, the refinement mechanism successfully synthesizes certificates for an additional 14 problems beyond the initial 49, bringing I_1 success to 63 solved problems. This demonstrates that even when the initial LLM-generated template requires adjustment, the refinement mechanism can improve coefficients and, when necessary, modify mathematical structures to satisfy all safety conditions.

Multi-Iteration Synthesis Impact. As demonstrated in Table 3, for problems not solved in the first iteration, the framework’s ability to generate structurally different templates in later iterations proves essential, bringing additional 27 successful barrier synthesis. The second iteration (I_2) successfully solves 15 additional problems (16.7%), bringing the cumulative success rate to 86.7%. The remaining 12 problems (13.3%) require three or more iterations (I_{3+}), with half of these requiring structural refinement (R_{3-4}), indicating that more challenging problems often require exploration of alternative mathematical forms, as the proportion of problems requiring structural refinement increases in later iterations (from 7.9% in I_1 to 50% in I_{3+}). The results of the ablation study provide a positive answer to **RQ3**, demonstrating the improvement in the efficiency and reliability of LLM-guided barrier synthesis through the agents interaction.

Limitations. Although our synthesis framework outperforms the baselines by solving 90% of the BARRIERBENCH problems, the remaining 10 unsolved problems reveal some synthesis challenges. The unsolved cases are highly complex nonlinear dynamics, including systems with complex combination of trigonometric and exponential terms, across different topological sets. Their complexity challenges both the synthesis framework and the verification solvers. In certain cases, the Synthesis Agent fails to generate barrier certificates satisfying all safety conditions due to the complexity of the underlying dynamics. Moreover, for the synthesized barriers, SMT solvers face computational difficulties during verifying conditions as system dimension and nonlinear complexity increase, frequently timing out or failing to determine satisfiability.

5. Related Work

Barrier certificate generation. Classical approaches to barrier certificate synthesis are predominantly optimization-based, especially sum-of-squares (SOS) programming (Prajna et al., 2007; Ames et al., 2019; Wooding et al., 2026; Kordabad et al., 2025), but their computational cost grows quickly with system dimension and complexity. To improve scalability, counterexample-guided inductive synthesis (CEGIS) frameworks such as FOSSIL (Abate et al., 2021) integrate neural networks with SMT-based verification, enabling iterative refinement of nonlinear barrier certificates beyond fixed polynomial templates. Barrier certificates have also been extended beyond safety to richer specifications, including signal temporal logic (Lindemann and Dimarogonas, 2019), linear temporal logic over finite traces (Jagtap et al., 2021), and reachability specifications (Majumdar et al., 2024).

When system dynamics are difficult to model exactly, data-driven approaches aim to infer safe behavior from sampled trajectories (Nejati et al., 2023). Much of this literature focuses on linear or control-affine dynamics (Jagtap et al., 2020; Cohen et al., 2022; Lopez and Slotine, 2023), often under partial knowledge or structural assumptions needed for formal guarantees. For example, (Wang et al., 2018) learns unknown dynamics with Gaussian processes while assuming known control-affine structure, and (Jagtap et al., 2020) assumes known control input while learning the unknown drift. Other works impose assumptions on stochastic components, for example by assuming

unknown noise distributions over known deterministic dynamics (Mathiesen et al., 2024) or known noise distributions (Mazouz et al., 2024). More recent work considers systems with fully unknown dynamics (Salamati et al., 2024; Wang et al., 2023; Schön et al., 2024), supported by tools such as LUCID (Casablanca et al., 2026). Despite these advances, most existing methods still rely on manually chosen templates, motivating more flexible and automated synthesis approaches.

LLMs for complex reasoning. Recent LLMs exhibit strong reasoning, planning, and problem-solving abilities (Achiam et al., 2023; Xu et al., 2024a; Zhao et al., 2023). Models such as GPT-3 (Brown et al., 2020; Dale, 2021) and GPT-4 (Sanderson, 2023) demonstrate broad generalization grounded in large-scale pretraining (Xu et al., 2024a). They can generate multi-step plans from high-level instructions (Jansen, 2020), decompose complex problems into subproblems, and reason across multiple levels of abstraction (Lin et al., 2023; Rana et al., 2023). These capabilities have been applied in domains ranging from robotic task planning (Rana et al., 2023) and autonomous driving (Cui et al., 2024) to mathematical reasoning involving symbolic manipulation, quantitative analysis, and formal logic (Lewkowycz et al., 2022). This progress makes LLMs promising tools for safety-critical settings that require structured reasoning about system constraints and safe behavior.

LLMs for CPS. LLMs are increasingly being explored in CPS as engines for high-level reasoning, planning, and natural-language interaction (Xu et al., 2024a). Their broad pretraining across language, code, mathematics, and science gives them a wide base of transferable knowledge. In robotics, SayCan (Ichter et al., 2023) combines LLM reasoning with reinforcement learning by using affordance-based value functions to ground language commands in executable actions. In autonomous driving, DriveGPT4 (Xu et al., 2024b) uses LLMs for interpretable decision-making, generating natural-language explanations together with low-level control predictions. In formal controller synthesis, (Bayat et al., 2025) propose a multi-agent framework that translates natural-language specifications into verified symbolic control code by coupling reasoning with abstraction-based verification. At the same time, LLMs remain prone to hallucination (Xu et al., 2024a), lack intrinsic grounding in physical constraints, and are difficult to verify with traditional neural-network verification tools such as Reluplex. These limitations motivate hybrid frameworks that combine LLM-based reasoning with formal verification to ensure correctness and safety in CPS.

6. Conclusion

We presented an LLM-agentic framework for barrier certificate synthesis that combines natural-language reasoning, retrieval of relevant prior barriers, and verification-guided refinement to address key limitations of existing methods. The framework automates much of the manual process of template selection, discovers diverse barrier structures beyond fixed-template approaches, and extends to barrier-controller co-synthesis for controlled systems. We also introduced BARRIERBENCH, a benchmark of 100 dynamical systems spanning continuous-time, discrete-time, linear, nonlinear, and controlled settings, on which the framework achieves a 90% success rate.

This work takes an initial step toward integrating language-based reasoning with formal safety verification and control synthesis. Rather than claiming exhaustive coverage, we aim to provide a concrete and extensible starting point for future work. We hope that releasing both the framework and BARRIERBENCH will support shared experimentation and accelerate progress toward more general, interpretable, and automated methods for safety assurance in dynamical systems.

Acknowledgements

This research was supported in part by the National Science Foundation under CAREER Award CCF-2146563. Ashutosh Trivedi is a Royal Society Wolfson Visiting Fellow and gratefully acknowledges support from the Wolfson Foundation and the Royal Society. The research of Sadegh Soudjani is supported by ERC 101089047 and by EIC 101070802. We also acknowledge the use of the Excalidraw tool (<https://excalidraw.com>) in creating the illustration in Figure 1.

References

- Alessandro Abate, Daniele Ahmed, Alec Edwards, Mirco Giacobbe, and Andrea Peruffo. FOSSIL: a software tool for the formal synthesis of Lyapunov functions and barrier certificates using neural networks. In *Proceedings of the 24th International Conference on Hybrid Systems: Computation and Control*, HSCC'21, New York, NY, USA, 2021. Association for Computing Machinery.
- Alessandro Abate, Sergiy Bogomolov, Alec Edwards, Kostiantyn Potomkin, Sadegh Soudjani, and Paolo Zuliani. Safe reach set computation via neural barrier certificates. *IFAC-PapersOnLine*, 58(11):107–114, 2024. 8th IFAC Conference on Analysis and Design of Hybrid Systems ADHS 2024.
- Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. GPT-4 technical report, 2023.
- Aaron D. Ames, Samuel Coogan, Magnus Egerstedt, Gennaro Notomista, Koushil Sreenath, and Paulo Tabuada. Control barrier functions: Theory and applications. In *2019 18th European Control Conference (ECC)*, pages 3420–3431, 2019.
- Eugene Asarin and Oded Maler. Achilles and the tortoise climbing up the arithmetical hierarchy. *Journal of Computer and System Sciences*, 57(3):389–398, 1998.
- Haniel Barbosa, Clark Barrett, Martin Brain, Gereon Kremer, Hanna Lachnitt, Makai Mann, Abdalrhman Mohamed, Mudathir Mohamed, Aina Niemetz, Andres Nötzli, Alex Ozdemir, Mathias Preiner, Andrew Reynolds, et al. cvc5: A versatile and industrial-strength SMT solver. In *Tools and Algorithms for the Construction and Analysis of Systems*, pages 415–442, Cham, 2022. Springer International Publishing.
- Amir Bayat, Alessandro Abate, Necmiye Ozay, and Raphaël M. Jungers. LLM-enhanced symbolic control for safety-critical applications. *IFAC-PapersOnLine*, 59(26):43–48, 2025. 7th IFAC Conference on Intelligent Control and Automation Sciences ICONS 2025.
- Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, et al. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc., 2020.
- Ernesto Casablanca, Oliver Schön, Paolo Zuliani, and Sadegh Soudjani. LUCID: Learning-enabled uncertainty-aware certification of stochastic dynamical systems. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(24):19916–19924, Mar. 2026.

- Max H. Cohen, Calin Belta, and Roberto Tron. Robust control barrier functions for nonlinear control systems with uncertainty: A duality-based approach. In *2022 IEEE 61st Conference on Decision and Control (CDC)*, pages 174–179, 2022.
- Can Cui, Zichong Yang, Yupeng Zhou, Yunsheng Ma, Juanwu Lu, Lingxi Li, Yaobin Chen, Jitesh Panchal, and Ziran Wang. Personalized autonomous driving with large language models: Field experiments. In *2024 IEEE 27th International Conference on Intelligent Transportation Systems (ITSC)*, pages 20–27. IEEE, 2024.
- Robert Dale. GPT-3: What’s it good for? *Natural Language Engineering*, 27(1):113–118, 2021.
- Charles Dawson, Sicun Gao, and Chuchu Fan. Safe control with learned certificates: A survey of neural Lyapunov, barrier, and contraction methods for robotics and control. *IEEE Transactions on Robotics*, 39(3):1749–1767, 2023.
- Leonardo de Moura and Nikolaj Bjørner. Z3: An efficient SMT solver. In *Tools and Algorithms for the Construction and Analysis of Systems*, pages 337–340. Springer Berlin Heidelberg, 2008.
- Bruno Dutertre. Yices 2.2. In Armin Biere and Roderick Bloem, editors, *Computer Aided Verification*, pages 737–744, Cham, 2014. Springer International Publishing.
- Brian Ichter, Anthony Brohan, Yevgen Chebotar, Chelsea Finn, Karol Hausman, Alexander Herzog, Daniel Ho, Julian Ibarz, Alex Irpan, Eric Jang, Ryan Julian, et al. Do As I Can, Not As I Say: Grounding language in robotic affordances. In *Proceedings of The 6th Conference on Robot Learning*, volume 205 of *Proceedings of Machine Learning Research*, pages 287–318. PMLR, 14–18 Dec 2023.
- Pushpak Jagtap, George J. Pappas, and Majid Zamani. Control barrier functions for unknown nonlinear systems using Gaussian processes. In *Proceedings of the IEEE Conference on Decision and Control*, volume 2020-Decem, pages 3699–3704, oct 2020.
- Pushpak Jagtap, Sadegh Soudjani, and Majid Zamani. Formal synthesis of stochastic systems via control barrier certificates. *IEEE Transactions on Automatic Control*, 66(7):3097–3110, 2021.
- Peter Jansen. Visually-grounded planning without vision: Language models infer detailed plans from high-level instructions. In *Findings of the association for computational linguistics: Emnlp 2020*, pages 4412–4417, 2020.
- Hassan K Khalil and Jessy W Grizzle. *Nonlinear systems*, volume 3. Prentice hall Upper Saddle River, NJ, 2002.
- Hui Kong, Fei He, Xiaoyu Song, William N. N. Hung, and Ming Gu. Exponential-condition-based barrier certificate generation for safety verification of hybrid systems. In *Computer Aided Verification*, pages 242–257, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.
- Arash Bahari Kordabad, Rupak Majumdar, and Sadegh Soudjani. Sum-of-squares certificates for almost-sure reachability of stochastic polynomial systems. *arXiv preprint arXiv:2510.25513*, 2025.

- Aitor Lewkowycz, Anders Andreassen, David Dohan, Ethan Dyer, Henryk Michalewski, Vinay Ramasesh, Ambrose Slone, Cem Anil, Imanol Schlag, Theo Gutman-Solo, Yuhuai Wu, et al. Solving quantitative reasoning problems with language models. In *Advances in Neural Information Processing Systems*, volume 35, pages 3843–3857. Curran Associates, Inc., 2022.
- Kevin Lin, Christopher Agia, Toki Migimatsu, Marco Pavone, and Jeannette Bohg. Text2Motion: from natural language instructions to feasible plans. *Auton. Robots*, 47(8):1345–1365, November 2023.
- Lars Lindemann and Dimos V. Dimarogonas. Control barrier functions for signal temporal logic tasks. *IEEE Control Systems Letters*, 3(1):96–101, 2019.
- Brett T. Lopez and Jean-Jacques E. Slotine. Unmatched control barrier functions: Certainty equivalence adaptive safety. In *2023 American Control Conference (ACC)*, pages 3662–3668, 2023.
- A. M. Lyapunov. *Stability of Motion*. Academic Press, London, 1966. Translated from the Russian edition, 1892.
- Rupak Majumdar, V. R. Sathiyarayanan, and Sadegh Soudjani. Necessary and sufficient certificates for almost sure reachability. *IEEE Control Systems Letters*, 8:2703–2708, 2024.
- Frederik Baymler Mathiesen, Licio Romao, Simeon C Calvert, Luca Laurenti, and Alessandro Abate. A data-driven approach for safety quantification of non-linear stochastic systems with unknown additive noise distribution. *arXiv preprint arXiv:2410.06662*, 2024.
- Rayan Mazouz, John Skovbeek, Frederik Baymler Mathiesen, Eric Frew, Luca Laurenti, and Morteza Lahijanian. Data-driven permissible safe control with barrier certificates. In *2024 IEEE 63rd Conference on Decision and Control (CDC)*, pages 6844–6849, 2024.
- Ameneh Nejati, Abolfazl Lavaei, Pushpak Jagtap, Sadegh Soudjani, and Majid Zamani. Formal verification of unknown discrete- and continuous-time systems: A data-driven approach. *IEEE Transactions on Automatic Control*, 68(5):3011–3024, 2023.
- Grigory Neustroev, Mirco Giacobbe, and Anna Lukina. Neural continuous-time supermartingale certificates. In *Proceedings of the Conf. on Artificial Intelligence, AAAI’25/IAAI’25/EAAI’25*. AAAI Press, 2025.
- Stephen Prajna and Ali Jadbabaie. Safety verification of hybrid systems using barrier certificates. In *Hybrid Systems: Computation and Control*, pages 477–492. Springer Berlin Heidelberg, 2004.
- Stephen Prajna, Ali Jadbabaie, and George J. Pappas. A framework for worst-case and stochastic safety verification using barrier certificates. *IEEE Transactions on Automatic Control*, 52(8):1415–1428, 2007.
- Krishan Rana, Jesse Haviland, Sourav Garg, Jad Abou-Chakra, Ian Reid, and Niko Suenderhauf. SayPlan: Grounding large language models using 3d scene graphs for scalable task planning. In *7th Annual Conference on Robot Learning*, 2023.

- Hadi Ravanbakhsh and Sriram Sankaranarayanan. Counter-example guided synthesis of control Lyapunov functions for switched systems. In *2015 54th IEEE Conference on Decision and Control (CDC)*, pages 4232–4239, 2015.
- Ali Salamati, Abolfazl Lavaei, Sadegh Soudjani, and Majid Zamani. Data-driven verification and synthesis of stochastic systems via barrier certificates. *Automatica*, 159:111323, 2024.
- Katharine Sanderson. GPT-4 is here: what scientists think. *Nature*, 615(7954):773, 2023.
- Oliver Schön, Zhengang Zhong, and Sadegh Soudjani. Data-driven distributionally robust safety verification using barrier certificates and conditional mean embeddings. In *2024 American Control Conference (ACC)*, pages 3417–3423, 2024.
- Chuanzheng Wang, Yiming Meng, Jun Liu, and Stephen Smith. Stochastic control barrier functions with Bayesian inference for unknown stochastic differential equations, 2023.
- Li Wang, Evangelos A. Theodorou, and Magnus Egerstedt. Safe learning of quadrotor dynamics using barrier certificates. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2460–2465, 2018.
- Ben Wooding, Viacheslav Horbanov, and Abolfazl Lavaei. PRoTECT: Parallelized construction of safety barrier certificates for nonlinear polynomial systems. In *Theoretical Aspects of Computing – ICTAC 2025*, pages 448–458, Cham, 2026. Springer Nature Switzerland.
- Weizhe Xu, Mengyu Liu, Oleg Sokolsky, Insup Lee, and Fanxin Kong. LLM-enabled cyber-physical systems: Survey, research opportunities, and challenges. In *2024 IEEE International Workshop on Foundation Models for Cyber-Physical Systems & Internet of Things (FMSys)*, pages 50–55, 2024a.
- Zhenhua Xu, Yujia Zhang, Enze Xie, Zhen Zhao, Yong Guo, Kwan-Yee K Wong, Zhenguo Li, and Hengshuang Zhao. DriveGPT4: Interpretable end-to-end autonomous driving via large language model. *IEEE Robotics and Automation Letters*, 9(10):8186–8193, 2024b.
- Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, et al. A survey of large language models, 2023.

Appendix A. LLM Prompts

This appendix presents the complete set of prompts used in our framework.

A.1. Prompt 1: Dataset Similarity Selection

This prompt selects the most similar problem from filtered candidates using LLM-based similarity assessment.

Dataset Similarity Selection

TARGET PROBLEM:

Dynamics: {SYSTEMDYNAMICS}
 Initial set: {INITIALSET}
 Unsafe set: {UNSAFE.SET}

COMPATIBLE CANDIDATES (all are fundamentally similar):

{CANDIDATES.TEXT}

Which candidate has the most similar problem type and structure to the target problem?

Focus on: system structure, problem type, and mathematical pattern similarity.

Answer with only the candidate number (1, 2, 3, etc.):

A.2. Prompt 2: Barrier Generation in the First Iteration

This prompt generates the barrier certificate for systems incorporating the retrieved context if available.

A.2.1. SYSTEMS WITHOUT CONTROLLER

Barrier Synthesis in the First Iteration - No Controller

{CONTEXT}

Main Problem:

- Dynamics: {SYSTEMDYNAMICS}
 - Initial set: {INITIALSET}
 - Unsafe set: {UNSAFE.SET}

Design a barrier certificate $B(x)$ that satisfies:

1. $B(x) \leq 0$ in initial set
2. $B(x) > 0$ in unsafe set
3. {CONDITION_3}

Be very careful - don't make it more complex than needed.

CRITICAL: Use ONLY real numbers in the barrier expression. No variables like 'c' or ' ϵ '.

Solve specifically for THIS problem with appropriate coefficients.

Analyze carefully but be concise. Give precise answer without long explanations.

Format your response as (don't make it bold):

BARRIER: [expression with numbers only]

The variable `CONDITION_3` refers to

- Discrete-time: $B(f(x)) - B(x) \leq 0$ for all x in the state space;
- Continuous-time: $\nabla B(x) \cdot f(x) < 0$ on the boundary.

The variable `CONTEXT` refers to the following two cases:

(a) when a similar problem exists in the dataset:

Related problem found:
 EXAMPLE: {SIMILAR_PROBLEM}
 B(x): {SIMILAR_BARRIER}

WARNING: This is just an example - your solution may be completely different NOT ONLY in terms of coefficients, BUT ALSO in format and structure. You may make mistakes, so do not consider this as a solution. Please don't copy it.

(b) when no similar problem exists:

No similar problems found. Analyze this problem fresh.

A.2.2. SYSTEMS WITH CONTROL INPUTS

Barrier and Controller Synthesis in the First Iteration

{CONTEXT}

Main Problem:

- Dynamics: {SYSTEM_DYNAMICS}
- Initial set: {INITIAL_SET}
- Unsafe set: {UNSAFE_SET}

CONTROLLER SYNTHESIS: This system has control inputs {CONTROLLER_PARAMETERS}.

You need to design BOTH:

1. Barrier certificate $B(x)$
2. Controller expressions for {CONTROLLER_PARAMETERS}

The controller $u(x)$ will be substituted into dynamics to create closed-loop system.

Design a barrier certificate $B(x)$ that satisfies:

1. $B(x) \leq 0$ in initial set
2. $B(x) > 0$ in unsafe set
3. {CONDITION_3}

Be very careful - don't make it more complex than needed.

Design both barrier certificate $B(x)$ AND controller expressions that work together to satisfy all conditions.

CRITICAL:

- Use ONLY real numbers in both barrier and controller expressions. No variables like 'c' or 'ε'.
- Solve specifically for THIS problem with appropriate coefficients.

- Controller must be implementable with realistic actuators
- Ensure controller bounds are reasonable (avoid extremely large values)

Analyze carefully but be concise. Give precise answer without long explanations.

Format your response as (don't make it bold):

BARRIER: [barrier expression with numbers only]

CONTROLLER: [controller expressions for each parameter, comma-separated]

A.3. Prompt 4: Barrier Generation in Subsequent Iterations

For subsequent iterations, the prompt leverages failure history to improve structure.

A.3.1. SYSTEMS WITHOUT CONTROLLER

Barrier Synthesis in Subsequent Iterations - No Controller

Previous attempts failed:

- Tried: {BARRIER} {FAILED_INFO}
- Tried: {BARRIER} {FAILED_INFO}
-

Improve the barrier structure to satisfy all conditions.

Main Problem:

- Dynamics: {SYSTEM_DYNAMICS}
- Initial set: {INITIAL_SET}
- Unsafe set: {UNSAFE_SET}

Design a barrier certificate $B(x)$ that satisfies:

1. $B(x) \leq 0$ in initial set
2. $B(x) > 0$ in unsafe set
3. {CONDITION_3}

Learn from previous failures. You can change structure of TEMPLATE if needed. In this step, the goal is to improve the structure of the templates, not refine the parameters.

CRITICAL: Use ONLY real numbers in the barrier expression. No variables like 'c' or ' ϵ '. Solve specifically for THIS problem with appropriate coefficients.

Analyze carefully but be concise. Give precise answer without long explanations.

Format your response as (don't make it bold):

BARRIER: [expression with numbers only]

The variable FAILED_INFO refers to

failed: FAILED_CONDITIONS (number of counter-examples).

A.3.2. SYSTEMS WITH CONTROL INPUTS

Barrier and Controller Synthesis in Subsequent Iterations

Previous barrier + controller attempts failed:

- Barrier: {BARRIER}, Controller: {CONTROLLER}, {FAILED_INFO}
- Barrier: {BARRIER}, Controller: {CONTROLLER}, {FAILED_INFO}
- Barrier: {BARRIER}, Controller: {CONTROLLER}, {FAILED_INFO}

Improve the barrier + controller structure to satisfy all conditions.

Main Problem:

- Dynamics: {SYSTEM_DYNAMICS}
- Initial set: {INITIAL_SET}
- Unsafe set: {UNSAFE_SET}

CONTROLLER SYNTHESIS:

This system has control inputs {CONTROLLER_PARAMETERS}. You need to design BOTH:

1. Barrier certificate $B(x)$
2. Controller expressions for {CONTROLLER_PARAMETERS}

The controller $u(x)$ will be substituted into dynamics to create closed-loop system.

Design a barrier certificate $B(x)$ that satisfies:

1. $B(x) \leq 0$ in initial set
2. $B(x) > 0$ in unsafe set
3. {CONDITION_3}

Learn from previous failures. You can change structure of TEMPLATE if needed. In this step, the goal is to improve the structure of the templates, not refine the parameters.

Design both barrier certificate $B(x)$ AND controller expressions that work together to satisfy all conditions.

CRITICAL:

- Use ONLY real numbers in both barrier and controller expressions. No variables like 'c' or 'ε'.
- Solve specifically for THIS problem with appropriate coefficients.
- Controller must be implementable with realistic actuators
- Ensure controller bounds are reasonable (avoid extremely large values)

Analyze carefully but be concise. Give precise answer without long explanations.

Format your response as (don't make it bold):

BARRIER: [barrier expression with numbers only]

CONTROLLER: [controller expressions for each parameter, comma-separated]

A.4. Prompt 6: Coefficient Refinement in the First Two Iterations

For the first two iterations of the refinement, the prompt focuses on the adjustment of the coefficient without structural changes.

A.4.1. SYSTEMS WITHOUT CONTROLLER

Coefficient Refinement - No Controller

Original barrier: {BARRIER} {FAILED_INFO}
 {REFINEMENT_HISTORY}

Problem:

- Dynamics: {SYSTEM_DYNAMICS}
- Initial set: {INITIAL_SET}
- Unsafe set: {UNSAFE_SET}

Try a different coefficient distribution. You can redistribute the coefficients between ALL terms (sometimes it is necessary for all terms to have different coefficients), but DO NOT change structure

Requirements:

1. $B(x) \leq 0$ in initial set
2. $B(x) > 0$ in unsafe set
3. {CONDITION_3}

Analyze carefully but be concise. Give precise answer without long explanations.

Format your response as (don't make it bold):

REFINED_BARRIER: [expression with numbers only]

A.4.2. SYSTEMS WITH CONTROL INPUTS

Coefficient Refinement - Barrier and Controller

Original barrier: {BARRIER}, Original controller: {CONTROLLER},
 {FAILED_INFO}
 {REFINEMENT_HISTORY}

Problem:

- Dynamics: {SYSTEM_DYNAMICS}
- Initial set: {INITIAL_SET}
- Unsafe set: {UNSAFE_SET}

Try a different coefficient distribution for both barrier and controller. You can redistribute the coefficients between ALL terms, but DO NOT change structure

CONTROLLER SYNTHESIS CONSTRAINTS:

1. Controller parameters: {CONTROLLER_PARAMETERS}
2. Use smooth, bounded functions (avoid extremely large values)
3. Controller must work harmoniously with the barrier
4. Ensure closed-loop stability

Requirements:

1. $B(x) \leq 0$ in initial set
2. $B(x) > 0$ in unsafe set
3. {CONDITION_3}

Analyze carefully but be concise. Give precise answer without long explanations.

Format your response as (don't make it bold):

REFINED_BARRIER: [barrier expression with numbers only]
 REFINED_CONTROLLER: [controller expressions for each parameter,
 comma-separated]

A.5. Prompt 7: Structure Refinement in Subsequent Iterations

For refinement iterations 3 and 4, the prompt allows structural modifications when coefficient adjustments fail.

A.5.1. SYSTEMS WITHOUT CONTROLLER

Structure Refinement - No Controller

Original barrier: {BARRIER} {FAILED_INFO}
 {REFINEMENT_HISTORY}

Problem:

- Dynamics: {SYSTEM_DYNAMICS}
- Initial set: {INITIAL_SET}
- Unsafe set: {UNSAFE_SET}

Previous coefficient adjustments failed. Consider changing the barrier structure if needed while keeping the same polynomial degree. You can modify the terms or their combinations.

Requirements:

1. $B(x) \leq 0$ in initial set
2. $B(x) > 0$ in unsafe set
3. {CONDITION_3}

Analyze carefully but be concise. Give precise answer without long explanations.

Format your response as (don't make it bold):

REFINED_BARRIER: [expression with numbers only]

A.5.2. SYSTEMS WITH CONTROL INPUT

Structure Refinement - Barrier and Controller

Original barrier: {BARRIER}, Original controller: {CONTROLLER},
 {FAILED_INFO}
 {REFINEMENT_HISTORY}

Problem:

- Dynamics: {SYSTEM_DYNAMICS}
- Initial set: {INITIAL_SET}
- Unsafe set: {UNSAFE_SET}

Previous coefficient adjustments failed. Consider changing the barrier and/or controller structure if needed while keeping the same polynomial degree. You can modify the terms or their combinations.

CONTROLLER SYNTHESIS CONSTRAINTS:

1. Controller parameters: {CONTROLLER_PARAMETERS}
2. Use smooth, bounded functions (avoid extremely large values)
3. Controller must work harmoniously with the barrier
4. Ensure closed-loop stability

Requirements:

1. $B(x) \leq 0$ in initial set
2. $B(x) > 0$ in unsafe set
3. {CONDITION_3}

Analyze carefully but be concise. Give precise answer without long explanations.

Format your response as (don't make it bold):

REFINED_BARRIER: [barrier expression with numbers only]

REFINED_CONTROLLER: [controller expressions for each parameter, comma-separated]

Refinement History Format:

```
Refinement 1: {BARRIER} {FAILED_INFO}
Refinement 2: {BARRIER} {FAILED_INFO}
...
```

A.6. Prompt 8: SMT Solver

A.6.1. SOLVER SELECTION

This prompt selects the most appropriate SMT solver for verification.

SMT Solver Selection

Select the best SMT solver based on barrier expression and the dynamical system.

PROBLEM:

Dynamics: {SYSTEM_DYNAMICS}

Barrier: {BARRIER_EXPRESSION}

AVAILABLE SOLVERS:

- cvc5
- z3
- yices

Without long explanations, give a short and precise answer.

Format your response as:

SOLVER: [solver name]

A.6.2. TIMEOUT ANALYSIS

This prompt determines whether to retry with extended timeout when a solver times out.

SMT Solver Timeout Analysis

Solver {SOLVER_NAME} timed out after {TIMEOUT_MS}ms.

PROBLEM:

Dynamics: {SYSTEM_DYNAMICS}

Barrier: {BARRIER_EXPRESSION}

Given the above information, should we attempt again with more time, or is this barrier too complex to verify?

Without long explanations, give a short and precise answer.

Format your response as:

RETRY: yes or no

TIMEOUT_MULTIPLIER: [number] (only if RETRY is yes, e.g., 1.5 or 2.0)

A.6.3. ERROR ANALYSIS

This prompt selects an alternative solver when verification fails.

Alternative Solver Selection

Solver {SOLVER_NAME} failed during verification.

ERROR: {ERROR_TYPE}

MESSAGE: {ERROR_MESSAGE}

PROBLEM:

Dynamics: {SYSTEM_DYNAMICS}

Barrier: {BARRIER_EXPRESSION}

REMAINING SOLVERS:

{REMAINING_SOLVERS_LIST}

Select a different solver to try.

Without long explanations, give a short and precise answer.

Format your response as:

NEXT_SOLVER: [solver name]