Orchestration of Dirac–Bergmann Hamiltonian analysis by large language models

Will Barker[1, *]

[1] *Central European Institute for Cosmology and Fundamental Physics, Institute of Physics of the Czech Academy of Sciences, Na Slovance 1999/2, 182 00 Prague 8, Czechia*

Control over the free-field particle spectrum is essential to constructing perturbative quantum field theories. For tensorial fields of any rank, and on flat spacetime, spectra are commonly obtained via spin-projection operators. This method does not readily extend to finite background fields, curved spacetime, or interactions beyond the free theory, for which the Dirac–Bergmann algorithm must instead be used. Whilst spin-projection reliably reduces to a linear algebra problem, the Dirac–Bergmann algorithm is more difficult to automate, for two reasons. Firstly, the computation and simplification of Poisson brackets poses a low-level problem that was only recently addressed with conventional computer algebra tools. Secondly, the flow of the algorithm depends on high-level decision making. Rather than resorting to a large and brittle decision tree, this second aspect might be easier to address in the natural language framework. We construct a fully autonomous multi-agent system that allows large language models to orchestrate the identification of all generations of constraints, their classification into those of first and second class, and the enumeration of physical degrees of freedom. The system relies on conventional computer algebra tools for the computation of Poisson brackets, and on inference for high-level decision making. The system is tested on a curated catalogue of models whose spectra are verified by spin-projection.

I INTRODUCTION

Importance of free-field spectra. — At the classical level, both gravity and the gauge sectors of the standard model can be built up perturbatively on flat spacetime by applying the Noether procedure to free field theories. In the former case, the massless Fierz–Pauli theory gives rise to general relativity [1–4], and in the latter the Maxwell theory gives rise to Yang–Mills [5]. Unitarity of the free theory remains essential even when quantum effects prohibit a perturbative treatment, and is usually taken as model-building criterion for developing putative new bosonic sectors. New bosonic sectors are of keen interest: they are a generic expectation of ultraviolet proposals [6–10], and appear in many specific scenarios beyond the standard model (see e.g. [11] or [12–14], or reviews [15, 16]).

Advantages of Spin projection. — Unitarity of the particle spectrum can be assessed for any freely-mixing configuration of tensorial fields, of any rank, by means of the spin-projection operator method [17–31] — see also [32, 33]. In this procedure, each field is decomposed into its irreducible parts under the action of the massive little group SO(3), and these spin components are arranged in the vector ψ_X. The wave operator in the k-space action $\mathscr{S}[\psi] \equiv \int \mathrm{d}^4k\ \psi^{*X}\mathsf{O}_X{}^Y\psi_Y$ is then expressed as a matrix $\mathsf{O}_X{}^Y$, for which any null vector $\mathsf{O}_X{}^Y\mathsf{v}_Y \equiv 0$ encodes the gauge symmetry $\delta\psi_X \propto \mathsf{v}_X$. Equally, if test sources are introduced as $\mathscr{S}[\psi]+\int \mathrm{d}^4k\ \psi^{*X}\mathsf{J}_X$, these null vectors enforce source constraints $\mathsf{v}^*{}_X\mathsf{J}^X = 0$. Bearing the null space in mind, the pseudoinverse $\mathsf{O}^{+X}{}_Y$ can be obtained; the sandwiched construction $\mathsf{J}^{*X}\mathsf{O}^+{}_X{}^Y\mathsf{J}_Y$ is the saturated propagator, from which the particle spectrum may be read off. The positions of the k-space propagator poles are square particle masses, whose positivity excludes tachyons; positivity of the pole residues excludes ghosts. In general, the residue is a quadratic form in the J^X components, whose rank indicates the number of polarizations (physical degrees of freedom) associated with the pole. So long as the little-group decomposition of all the fields is known, the spin-projection method can be applied. Since it reduces to a linear algebra problem, it will never fail, and moreover it can be easily implemented in computer algebra systems [34–39].

Disadvantages of spin projection. — If the little-group decomposition is inapplicable, the linear algebra problem cannot be so easily set up. This is generally the case when the isometry group changes, as in curved spacetime,[1] or when the fields have non-trivial background values which interfere with the local k-space representation. Moreover, interactions which extend $\mathscr{S}[\psi]$ to cubic and higher order in the fields deny a matrix formulation, while allowing mixing between spin sectors. In principle, the Noether and order-reduction procedures protect the spectrum from modifications due to interactions, but not all theories are motivated in this way. In particular, non-linear completions of *modified gravity* are frequently put forward as classical field theories. One way to rule these theories out is to check for strong coupling (see e.g. the well-known series [41, 42], and [43]), which manifests as sensitivity of the spectrum to the interactions, or to non-Minkowski backgrounds. The former are found by

* barker@fzu.cz

[1] See however [40] for successful extension to maximally-symmetric spacetimes.

Taylor-expanding the non-linear completion; the latter are phenomenologically desirable exact solutions to the modified field equations.

Advantages of Dirac–Bergmann. — In such cases, the Dirac–Bergmann algorithm offers a more general route to the spectrum [44–46] — see also [47–50]. In this prescription, the field components ψ_X and their conjugate momenta π^X are evolved between Cauchy surfaces according to first-order Hamilton equations, expressible in terms of Poisson brackets as $\dot{\psi}_X \approx \{\psi_X, \mathscr{H}\}$ and $\dot{\pi}^X \approx \{\pi^X, \mathscr{H}\}$, where $\mathscr{H}$ is the total Hamiltonian.[2] Notationally, we assume that ψ_X and π^X follow the same SO(3) decomposition of spin-projection, but this choice is inessential to the algorithm. In the definitions $\pi^X \equiv \delta\mathscr{S}[\psi]/\delta\dot{\psi}_X$, some velocities $\dot{\psi}_X$ may turn out to be insoluble in terms of the π^X and ψ_X. This gives rise to primary constraints $\phi^X \approx 0$, which are enforced in the position-space canonical action $\mathscr{S}[\psi,\pi,\lambda] \equiv \int \mathrm{d}t \ \left[\int \mathrm{d}^3x\, \pi^X\dot{\psi}_X - \mathscr{H}[\psi,\pi,\lambda]\right]$ by additional Lagrange multipliers λ_X such that $\mathscr{H}[\psi,\pi,\lambda] \supset \int \mathrm{d}^3x\, \lambda_X\phi^X$. For valid initial data, the $\phi^X \approx 0$ must be conserved in time. This requirement may give rise to secondary constraints $\chi^X \equiv \{\phi^X, \mathscr{H}\} \approx 0$, and so on, until each constraint chain terminates with either identical conservation or an on-shell solution for some λ_X. In total one may count N_{FC} first-class (FC) constraints whose Poisson brackets vanish with all the rest, and N_{SC} second-class (SC) constraints with mutually non-vanishing Poisson brackets. Given N_{Can} canonical components among the π^X and ψ_X, there are N_{Phy} physical modes in the spectrum, where

$$N_{\mathrm{Phy}} = \frac{1}{2}\Big(N_{\mathrm{Can}} - 2N_{\mathrm{FC}} - N_{\mathrm{SC}}\Big)\,. \tag{1}$$

As with spin-projection, the Dirac–Bergmann algorithm provides insight into the gauge structure, since FC $\phi^X \approx 0$ are identified as symmetry generators which allow the explicit transformations to be reconstructed by means of the Castellani algorithm [48].[3] Unlike for the case of spin-projection, the Dirac–Bergmann algorithm is applicable beyond free theories, since interactions may be included in $\mathscr{H}$ — the method is inherently non-perturbative.

Disadvantages of Dirac–Bergmann. — The formula in Eq. (1) does not inform us about the unitarity of the S-matrix, but this can usually be achieved for some volume of the space of coupling constants if N_{Phy} is a sufficiently small number.[4] A more serious problem is that, unlike for the spin-projection approach, the Dirac–Bergmann algorithm is resistant to automation. There are two reasons for this:

- At the lower level, the computation of Poisson brackets can be computationally expensive. This is especially true in the field theory context, where spatial gradients may necessitate many integrations by parts and dimensionally-dependent identities to simplify a bracket.
- At the higher level, the progress of the algorithm does not follow a rigid procedure. Decisions must be made about constraint independence, class, and conservation.

The problem of Poisson brackets is amenable to a brute-force computer algebra approach, and the software package *Hamilcar* was put forward in [53] to facilitate this based on pre-existing frameworks [54–57] and implementations [58]. This software was shown to be effective against bracket and constraint algebra reconstruction problems in the context of non-linear gravity [43, 59, 60]. On the other hand, the problem of automatic decision making remains unresolved.

Natural language orchestration. — Whilst the flow of the Dirac–Bergmann algorithm cannot be hard-coded without a large and brittle decision tree, it does follow set patterns that may be articulated in natural language. Various large language models (LLMs) have recently been enabled by empirical scaling-law [61] and architectural [62] advances, which show promise for following complex instructions in natural language, and for applying them in novel contexts. It is therefore interesting to consider whether LLMs may be employed to orchestrate the Dirac–Bergmann algorithm at the higher level, delegating the lower-level Poisson bracket computations to a computer algebra tool. This delegation is crucial: LLMs do *not* have an aptitude for calculations (they often fail to perform even basic arithmetic), and moreover remain susceptible to hallucinations. These problems can sometimes be ameliorated in multi-agent systems (MASs) which enforce tool use. Such systems are currently employed in precision cosmology [63–65], and have been proposed in other fields [66]. The Dirac–Bergmann algorithm is a specialised application in theoretical high-energy physics, where LLM capability is also under scrutiny [67, 68]. Successful automation would advance our understanding of LLM capabilities in theoretical physics, whilst also being inherently useful in that field.

In this paper. — We demonstrate a tool-augmented MAS operating entirely at inference time, which orchestrates the Dirac–Bergmann algorithm for free bosonic field

[2] Properly, the constraints in these Poisson brackets should be smeared; we introduce smearing notation in Section II A.

[3] Indeed, the FC algebra facilitates the quantisation of gauge theories [51, 52].

[4] The volume aspect is important. Unitarity conditions which specify some set-of-measure-zero hypersurface in the remaining coupling space (after the minimal tuning needed to obtain the gauge symmetry) are vulnerable to radiative corrections. For example, the symmetries of the Fierz–Pauli and Maxwell theories each leave a single global coupling: unitarity fixes the sign, not the value of this coupling, i.e. half the remaining volume.

theories. In principle, the MAS is model-agnostic, but it is tested here using *OpenAI*'s *GPT-5.2* (API identifier **`gpt-5.2`**) model [69, 70]. The MAS is described in Section II A, means for testing it in Section II B, and some examples of its use given in Section II C. Conclusions follow in Section III. We use natural units in which $\hbar = c = 1$. Greek letters denote four-dimensional spacetime indices, while Roman letters denote three-dimensional spatial indices. The signature is $(-,+,+,+)$, other conventions are introduced as needed.

II IMPLEMENTATION

A. Multi-agent system

Overview. — The structure of the MAS is illustrated in Fig. 1, and the sources are provided in the supplement [71]. The MAS is heterogeneous: it consists of a deliberative LLM agent which can interface via the model context protocol (MCP) with a computer algebra tool for computing Poisson brackets, and whose activity is regulated by a simple, fixed-policy read-evaluate-print-loop (REPL) agent.

LLM agent. — The LLM agent uses *OpenAI*'s *Agents SDK* to orchestrate the agentic loop: receiving input, invoking tools as needed, and iterating until the LLM produces a final response. Whilst *Agents SDK* is model-agnostic, our results are obtained using *OpenAI*'s *GPT-5.2* (API identifier **`gpt-5.2`**) [69, 70], announced in December 2025. This variant supports adaptive reasoning and performant tool-calling, with a 400,000 token context window (of which the system prompt is a generically small fraction). Model parameters are left as default values, except for **`reasoning_effort`** which is set to **`medium`**. Ideally, ample literature regarding the Dirac–Bergmann algorithm and *Wolfram Language* will feature in the LLM training corpus. The presence of this literature is verifiable for many LLMs [72], and while it cannot be confirmed in the case of **`gpt-5.2`** due to the closed nature of the model, it remains a reasonable assumption.

System prompt. — The system prompt is composed of three components:

1. Behaviour as natural language, encouraging:
   - Interaction with the REPL agent. At each turn the LLM agent is instructed to perform operations permitted by the REPL agent's previous response, articulate the results, and propose a limited next step in the algorithm. In case the LLM agent feels that the algorithm is terminated, it may include the string **`TERMINATE`** in its response. The LLM agent anticipates a **`yes`**/**`no`** response from the REPL agent at every turn except the first and last.
   - Suppression of overly-ambitious step-sizes.
   - Enhanced reasoning during certain steps.
   - Error-recovery strategies.
2. The *Hamilcar* sources as *Wolfram Language*, together with a navigation tree.
3. Two worked examples of obfuscated test models as *Wolfram Language*, with natural language commentary: Proca theory and Kalb–Ramond theory. The theories are deliberately simple, and have limited overlap in their constraint structure with the examples in Section II C.

In particular, the Dirac–Bergmann algorithm itself is not explained to the LLM as part of the system prompt, except through the two worked examples.

MCP server. — A persistent *Wolfram* (formerly known as *Mathematica*) kernel session maintains state across tool calls. The kernel connection is managed via *Wolfram Research*'s official **`wolframclient`** *Python* library, which provides a **`WolframLanguageSession`** class for inter-process communication. The MCP server exposes **`tool_WolframScript`** which allows the LLM to send arbitrary *Wolfram Language* code to the kernel. The system prompt enforces the use of the *Hamilcar* package for computing Poisson brackets [53], which is part of the *xAct* ecosystem [54–57]. Performance is not improved by providing specialized tools for specific *Hamilcar* functions: the model demonstrates a preference and aptitude for free composition.

REPL agent. — The REPL agent interprets LLM responses using regular expressions and a hard-coded decision tree, and serves several purposes:

- Step-size moderation via **`yes`**/**`no`** responses.
- Additional error-recovery strategies.
- Solicitation of a final summary.

Broadly, the REPL agent prevents the LLM agent from attempting overly-ambitious tasks. Meanwhile, it encourages continued progress to completion, excluding the user from the end-to-end flow of the Dirac–Bergmann algorithm. It is interesting that a regex script is sufficient to act as the second agent in an effective MAS. Improved per-turn performance may be expected from a second LLM agent, and indeed this would be less cumbersome to prompt than a REPL to script. The key point, however, is that a second LLM agent is *not necessary* for successfully automating the Dirac–Bergmann algorithm, while the script-based approach drives down inference costs.

B. Performance

Overview. — The viability of the MAS is tested against a curated catalogue of free-field theories whose spectra are known from independent applications of the spin-projection method, as discussed in Section I.

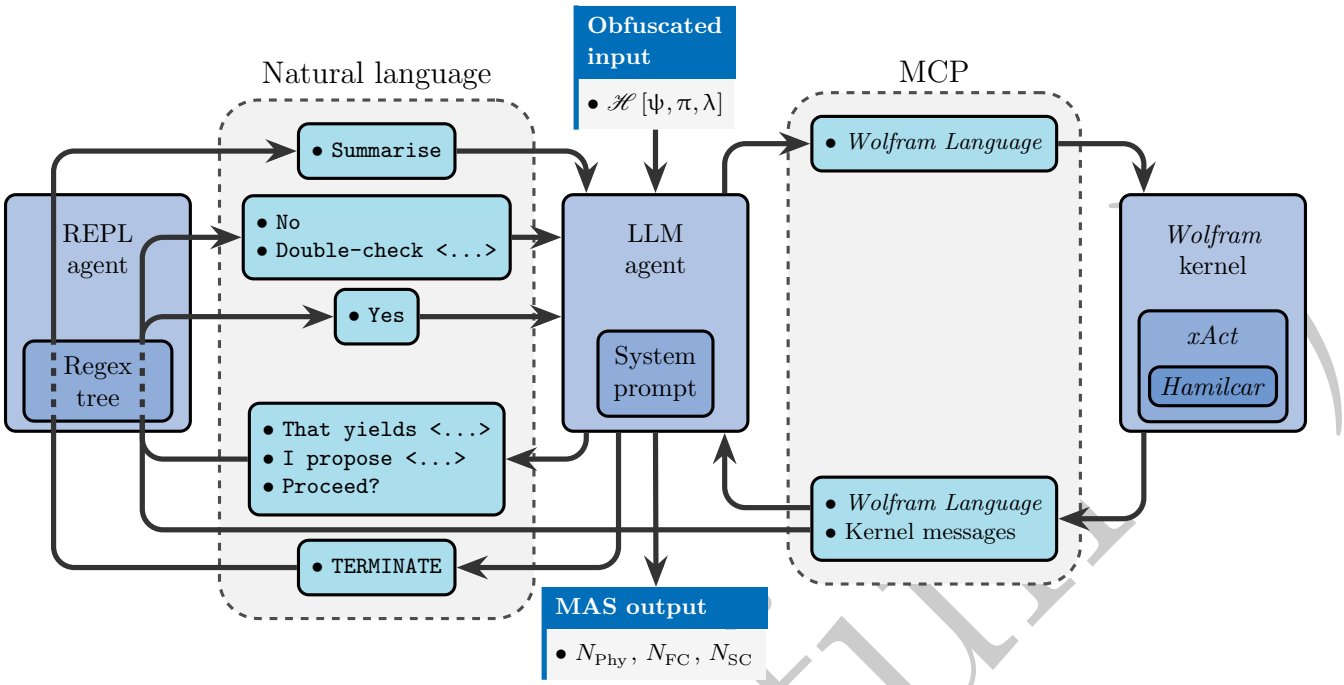

FIG. 1. Structure of the multi-agent system (MAS) that implements the Dirac–Bergmann algorithm from start to finish. The large language model (LLM) agent orchestrates the algorithm, and relies on tool calls via the model context protocol (MCP) to compute Poisson brackets. A read-evaluate-print-loop (REPL) agent acts as a regulator.

Independent tests by spin-projection. — For free theories, the results of the Dirac–Bergmann analysis — however protracted to implement — may be quickly obtained by the spin-projection method. For the latter, we employ the *PSALTer* software [38, 39] for *xAct* and *Wolfram*. The input to *PSALTer* is the action $\mathscr{S}[\psi]$ in which the fields are *not* yet decomposed into their ψ_X components. This decomposition is performed internally by means of SO(3) projection operators which are constructed from the Minkowski metric $\eta_{\mu\nu}$ and the unit-timelike vector n^μ. This unit-timelike vector has the property $n^\mu n_\mu \equiv -1$ and the interpretation that it be defined according to the massive particle four-momentum k^μ according to $n^\mu \equiv k^\mu/\sqrt{-k^2}$, where $k^2 \equiv k^\mu k_\mu$. The output of *PSALTer* is a particle spectrograph with rich details, of which we need only two pieces of information:

- The total number of polarizations among resolved and unresolved poles,[5] since this counts the number of physical modes N_{Phy}.
- The total number of independent components among the source constraints ${\mathsf{v}^*}_X \mathsf{J}^X = 0$, since this corresponds to the number of primary $\phi^X \approx 0$ among the N_{FC}.

[5] An unresolved pole is one whose position in k-space cannot be expressed in terms of rational functions of the couplings in $\mathscr{S}[\psi]$.

Extraction of the total Hamiltonian. — For free theories, $\mathscr{H}[\psi,\pi,\lambda]$ can be computed automatically using a script based on certain routines internal to the *PSALTer* codebase. This is made possible by the linearity of the theory and the properties of the SO(3) decomposition. The unit-timelike vector n^μ provides a natural global frame (the massive particle rest-frame) with which to define the foliation. The projected ψ_X are naturally orthogonal to n^μ in all their indices, so that once the inverse metric $\eta^{\mu\nu} = -\delta^{\mu\nu} + n^\mu n^\nu$ is separated in the position-space representation of $\mathscr{S}[\psi]$, the unit-timelike vector is everywhere confined to the scalar operator $n^\mu\partial_\mu$. It is then safe to identify the velocities $\dot{\psi}_X \equiv n^\mu\partial_\mu\psi_X$. The definitions $\pi^X \equiv \delta\mathscr{S}[\psi]/\delta\dot{\psi}_X$ constitute a linear system of equations that may be programmatically inverted — as far as possible — to express the $\dot{\psi}_X$ in terms of the π^X and ψ_X. The part of this system that does not admit any solution corresponds to the $\phi^X \approx 0$. Any dependence on $\dot{\psi}_X$ that does not cancel in the Legendre transformation may be substituted with the corresponding λ_X to yield $\mathscr{H}[\psi,\pi,\lambda]$, which corresponds to a shift of the λ_X that leaves the physics unchanged.

Model obfuscation. — Whilst the unavoidable presence of explicit examples in the training corpus is doubtless beneficial for the capabilities of the system, it causes a problem for assessing the MAS's flexibility. It is therefore important to remove explicit reference to specific theories from $\mathscr{H}[\psi,\pi,\lambda]$. To achieve this, the

fields ψ_X, momenta π^X, multipliers λ_X and coupling constants in $\mathscr{H}[\psi,\pi,\lambda]$ are systematically obfuscated by renaming them with hashes, while the model is always referred to anonymously in the natural language dressing. This approach at least places all the test theories on a maximally equal footing at inference time.

Failure modes and error recovery. — Given the MAS architecture explained in Section II A, several failure modes remain resistant to prompt engineering:

- Failure to identify spatial gradients of constraints as essentially constraining a reduced number of degrees of freedom. In such cases, constraint chains become artificially protracted, and the final constraints are over-counted. This becomes apparent to the LLM agent when $N_{\text{Phy}} < 0$. Recovery is hindered by the length of the turn sequence between the first error and the final counting. Such a correlation would be expected if, for example, the LLM agent tended to assume that the majority of the turn sequence must be correct when attempting to locate the source of an error.

- Failure to recognise that one constraint — as initially obtained — is linearly dependent on others. The consequences are as above.

- Failure to construct FC combinations of naïvely SC constraints. The consequences are as above, though in this case the error is signalled by $N_{\text{Phy}} = (n+1)/2$ for $n \in \mathbb{Z}$.

- Failure of *Wolfram Language* syntax. The LLM agent occasionally attempts to define `snake_case` variables: these conflict with pattern-matching in *Wolfram Language*, and it is to be noted that the *Hamilcar* sources and obfuscated models consistently employ rigid `PascalCase` naming conventions throughout. Recovery is usually immediate in such cases, since the kernel messages are unambiguous.

- Failure of *xAct* syntax. The consequences are as above. It is to be expected that the training corpus is starved of *xAct*-specific examples.

- Failure to distinguish between actual and proposed tool calls. To some extent, this appears to be an engineering failure specific to the *Agents SDK* infrastructure. Recovery is greatly improved by injecting natural language reminders of recent tool calls via the REPL agent.

None of the failure modes listed above reflect any fundamental incapacity of the **gpt-5.2** model, since all occur stochastically and infrequently, if at all, for repeat evaluations of the same obfuscated theory. The ease of recovery is a key difference between algorithmic implementations and agents, and is facilitated by access to a persistent kernel session.

C. Examples

Massive longitudinal vector. — We first consider a class of theories that is simple enough to plausibly have appeared in the training corpus of the LLM, specifically the vector theory in which there is a gauge invariance under transverse shifts of the vector. In the action

$$\mathscr{S} = \int \mathrm{d}^4x \left[-\frac{\alpha}{4}\partial_\mu A^\mu \partial_\nu A^\nu - \frac{\beta}{2} A_\mu A^\mu \right], \tag{2}$$

such a symmetry arises when $\beta = 0$. For the field A_μ the two elements of ψ_X are

$$A^{\#1}_{0^+} \equiv -n^\mu A_\mu\,, \quad A^{\#1}_{1^-\mu} \equiv \left(\delta^\nu_\mu + n_\mu n^\nu\right) A_\nu\,, \tag{3}$$

in the standard notation of [38, 39] where the subscript J^P describes the spin J and parity P of each mode, and the possibility of multiple modes with the same J^P is indicated by a superscript. In general, the spin-parity modes may carry Greek indices, but in such cases it is assumed that contraction of any Greek indices with n^μ yields zero. In the canonical picture we allow n^μ to define the foliation, so that in adapted coordinates it is sufficient to work with Roman indices only. This notation extends naturally to conjugate momenta $\pi(A)^{\#1}_{0^+}$ and $\pi(A)^{\#1^i}_{1^-}$, and Lagrange multipliers $\lambda(A)^{\#1}_{0^+}$ and $\lambda(A)^{\#1}_{1^-i}$. It is easy to see that the total Hamiltonian corresponding to Eq. (2) is

$$\begin{aligned}\mathscr{H} = \int \mathrm{d}^3x \Bigg[&-\frac{\pi(A)^{\#1\,2}_{0^+}}{\alpha} + \frac{\beta}{2} A^{\#1\,2}_{0^+} - \frac{\beta}{2} A^{\#1}_{1^-i} A^{\#1 i}_{1^-} \\ &+ \pi(A)^{\#1}_{1^-i} \lambda(A)^{\#1^i}_{1^-} + \pi(A)^{\#1}_{0^+} \partial_i A^{\#1^i}_{1^-} \Bigg],\end{aligned} \tag{4}$$

and that a single primary constraint arises as

$$\phi(A)^{\#1}_{1^-i} \equiv \pi(A)^{\#1}_{1^-i} \approx 0\,. \tag{5}$$

Conservation of the primary constraint yields[6]

$$\begin{aligned}\dot{\phi}(A)^{\#1}_{1^-i} &\approx \frac{\delta}{\delta f^i}\left\{\phi(A)^{\#1 j}_{1^-}[f_j], \mathscr{H}\right\} \\ &= \partial_i \pi(A)^{\#1}_{0^+} + \beta A^{\#1}_{1^-i} \approx 0\,.\end{aligned} \tag{6}$$

In the general case, this implies the secondary constraint

$$\chi(A)^{\#1}_{1^-i} \equiv \partial_i \pi(A)^{\#1}_{0^+} + \beta A^{\#1}_{1^-i} \approx 0\,. \tag{7}$$

The primary and secondary are found to be SC

$$\left\{\phi(A)^{\#1^i}_{1^-}[f_i], \chi(A)^{\#1^j}_{1^-}[s_j]\right\} = \beta[f_i s^i]\,, \tag{8}$$

which allows the multiplier $\lambda(A)^{\#1^i}_{1^-}$ to be determined. The total number of SC constraints is $N_{\text{SC}} = 6$, while the total

[6] We introduce smearing functions f_i and s_i; the notation is such that (e.g.) $\phi(A)^{\#1^i}_{1^-}[f_i] \equiv \int \mathrm{d}^3x\; f_i \phi(A)^{\#1^i}_{1^-}$.

number of canonical variables is $N_{\text{Can}} = 8$. Since there are no FC constraints, the final number of propagating modes is found from Eq. (1) to be

$$N_{\text{Phy}} = \frac{1}{2}\Big(8 - 6\Big) = 1\,. \tag{9}$$

Thus, the theory propagates a single massive scalar. In the spin-projection picture, the wave operator associated with Eq. (2) — split into its J^P blocks — is

$$\mathsf{O}_{0^+} = \begin{bmatrix} & A^{\#1}_{0^+} \\ A^{\#1\dagger}_{0^+} & \frac{1}{4}(-k^2\alpha - 2\beta) \end{bmatrix}, \tag{10a}$$

$$\mathsf{O}_{1^-} = \begin{bmatrix} & A^{\#1^i}_{1^-} \\ A^{\#1\dagger}_{1^-\,i} & -\frac{\beta}{2} \end{bmatrix}. \tag{10b}$$

From Eq. (10a) the single polarization with square mass $-2\beta/\alpha$ and pole residue proportional to α is visible. Component-wise analysis of Eqs. (10a) and (10b) in the lightcone limit does not reveal any 'hidden' massless sectors. The total Hamiltonian defined in Eq. (4) can be obfuscated as follows:

Obfuscated input

```
        ( CouplingConstante5f53ecd *
↪ CanonicalFieldccb6bb11[]^2 ) / 2 -
↪ ConjugateMomentumCanonicalFieldccb6bb11[]^2
↪ / CouplingConstant750dbd1c - (
↪ CouplingConstante5f53ecd *
↪ CanonicalField1ea68330[-a] *
↪ CanonicalField1ea68330[a] ) / 2 +
↪ ConjugateMomentumCanonicalField1ea68330[-a]
↪ * LagrangeMultiplier1ea68330[a] +
↪ ConjugateMomentumCanonicalFieldccb6bb11[]
↪ * CD[-a][CanonicalField1ea68330[a]]
```

Note that quantities which are pre-defined in *Hamilcar* are denoted in **BlueBold**; these include the partial derivative **CD[-i]@**, i.e. ∂_i, and all lower-case Roman indices **a**, **b**, **c**, etc., up to **z**, which are reserved for spatial indices. Quantities in SkyBlueBold are the obfuscated variables. In order to analyze the theory, the LLM agent is required to perform multiple operations on the state of the kernel session:

- Definition of the fields and their conjugate momenta using *Hamilcar*'s **DefCanonicalField** function.
- Definition of the Lagrange multipliers and smearing functions using *xAct*'s **DefTensor** command.
- Definition of the coupling constants using *xAct*'s **DefConstantSymbol** command.
- Identification of all primary constraints as variations with respect to Lagrange multipliers by using *xAct*'s **VarD** function.
- Computation of the smeared primary constraint velocities using *Hamilcar*'s **PoissonBracket** function.
- Subsequent steps using the above functions, in combinations deemed appropriate by the LLM agent, until termination. The agent must pay attention to the Einstein summation convention, symmetry of mixed partials, full list of intermediate variables and other aspects.

When this scripted $\mathscr{H}$ is presented to the MAS, the natural language response with the variables unblinded a posteriori is:

MAS output

This total Hamiltonian describes a theory on flat spatial slices with canonical variables $A^{\#1}_{0^+}$, $\pi(A)^{\#1}_{0^+}$ and $A^{\#1}_{1^-i}$, $\pi(A)^{\#1^i}_{1^-}$, and with constant symbols α and β. The Dirac–Bergmann analysis yields the following constraints in terms of canonical variables. The primary constraint is $\pi(A)^{\#1^i}_{1^-} \approx 0$, which is a spatial vector constraint and hence corresponds to 3 independent constrained degrees of freedom (one for each spatial component). Preservation in time of this primary constraint produces the secondary constraint $\beta A^{\#1^i}_{1^-} + \partial^i \pi(A)^{\#1}_{0^+} \approx 0$, also a spatial vector and hence likewise imposing 3 independent constrained degrees of freedom. No tertiary constraints arise.
For generic nonzero β, these two vector constraints are SC as a pair: their smeared Poisson bracket is nonvanishing and proportional to β, while the primary–primary and secondary–secondary brackets vanish. Consistency (time preservation) fixes the Lagrange multiplier $\lambda(A)^{\#1}_{1^-i}$ (all 3 of its components are determined by the consistency conditions). No other Lagrange multiplier is fixed by consistency; in particular, $\lambda(A)^{\#1}_{0^+}$ does not generate a constraint from the given total Hamiltonian and remains undetermined.
The total number of canonical degrees of freedom is 2 times the number of configuration components, i.e. $2 \times (1 + 3) = 8$ from the scalar $A^{\#1}_{0^+}$ and the spatial vector $A^{\#1}_{1^-i}$. With 6 SC constrained degrees of freedom in total (3 primary + 3 secondary) and no FC constraints, the number of physical degrees of freedom is $(8 - 6)/2 = 1$.

Thus, the MAS correctly reproduces the full chain of reasoning in Eqs. (5) to (9).

Massless longitudinal vector. — In the case that $\beta = 0$, Eq. (6) instead implies the secondary constraint

$$\chi(A)^{\#1}_{0^+} \equiv \pi(A)^{\#1}_{0^+} \approx 0\,. \tag{11}$$

It is obvious from Eq. (4) that the conservation in time of $\chi(A)^{\#1}_{0^+}$ is identically satisfied, and that $\phi(A)^{\#1}_{1^-i}$ and $\chi(A)^{\#1}_{0^+}$ commute and are FC. The total number of FC constraints is $N_{\text{FC}} = 4$, while the total number of canonical variables is $N_{\text{Can}} = 8$. Since there are no SC

constraints, the final number of propagating modes is found from Eq. (1) to be

$$N_{\rm Phy} = \frac{1}{2}\Big(8 - 2\times 4\Big) = 0\,. \tag{12}$$

This is a slightly surprising result, but the vanishing of all modes is due to the gauge symmetry which emerges in the massless limit, and which is generated by the primary FC constraint $\phi(A)^{\#1}_{1^-i}$. This is readily confirmed by the spin-projection approach: the matrix in Eq. (10) becomes singular and a component-wise analysis of the saturated propagator reveals that there is no pole, massless or otherwise. The $\beta = 0$ case of the total Hamiltonian defined in Eq. (4) can be obfuscated as follows:

Obfuscated input

```
        - (
↪ ConjugateMomentumCanonicalFieldccb6bb11[]^2
↪ / CouplingConstant750dbd1c ) +
↪ ConjugateMomentumCanonicalField1ea68330[-a]
↪ * LagrangeMultiplier1ea68330[a] +
↪ ConjugateMomentumCanonicalFieldccb6bb11[]
↪ * CD[-a][CanonicalField1ea68330[a]]
```

When this is presented to the MAS, the response is:

MAS output

This total Hamiltonian describes a theory with canonical variables $A^{\#1}_{0^+}$, $\pi(A)^{\#1}_{0^+}$, $A^{\#1}_{1^-i}$, and $\pi(A)^{\#1\,i}_{1^-}$ (with coupling α) and with the multiplier sector including $\lambda(A)^{\#1}_{0^+}$ and $\lambda(A)^{\#1}_{1^-i}$ as provided. The Dirac–Bergmann analysis yields no primary constraint from $\lambda(A)^{\#1}_{0^+}$ (its variational derivative of the total Hamiltonian density vanishes identically), while variation with respect to $\lambda(A)^{\#1}_{1^-i}$ yields the primary constraint $\pi(A)^{\#1\,i}_{1^-} \approx 0$, which is a rank-1 constraint with 3 independent components. Requiring time preservation of this primary constraint produces a secondary constraint $\partial^i\pi(A)^{\#1}_{0^+} \approx 0$; although written as a covector equation (3 components), it is a pure gradient of the scalar momentum $\pi(A)^{\#1}_{0^+}$ and therefore contributes 1 independent local constraint component (equivalently, it constrains $\pi(A)^{\#1}_{0^+}$ to vanish up to spatially constant modes, which are irrelevant for local degree-of-freedom counting). Preservation of the secondary constraint closes the algorithm with no tertiary constraints.
All Poisson brackets among the constraints vanish (in particular, the bracket between $\pi(A)^{\#1\,i}_{1^-}$ and $\partial^j\pi(A)^{\#1}_{0^+}$ vanishes), so both constraints are FC. No consistency condition determines the time evolution of any Lagrange multiplier: neither $\lambda(A)^{\#1}_{1^-i}$ nor $\lambda(A)^{\#1}_{0^+}$ is fixed by constraint preservation, and both remain undetermined.
The total number of canonical phase space degrees of freedom is $2\,(1+3) = 8$, corresponding to one scalar canonical pair $(A^{\#1}_{0^+}, \pi(A)^{\#1}_{0^+})$ and one spatial vector canonical pair $(A^{\#1}_{1^-i}, \pi(A)^{\#1\,i}_{1^-})$. The FC constraints remove 2×3 degrees of freedom from $\pi(A)^{\#1\,i}_{1^-} \approx 0$ plus 2×1 degrees of freedom from the independent content of $\partial^i\pi(A)^{\#1}_{0^+} \approx 0$, for a total reduction of $2\,(3+1) = 8$, leaving $8-8=0$ physical phase space degrees of freedom, i.e. 0 physical configuration-space degrees of freedom.

Thus, the MAS correctly reproduces the full chain of reasoning in Eqs. (11) to (12).

Massive multi-particle theory. — We now move beyond standard theories to consider models which are very unlikely to have appeared in the training corpus of the LLM. We consider the theory

$$\begin{aligned}\mathscr{S} = \int \mathrm{d}^4x \Big[&\alpha\partial_\mu A^\mu \partial_\nu A^\nu + \beta\partial_\mu B^\mu \partial_\nu B^\nu \\ &+ \gamma\partial_\mu B_\nu \partial^\mu B^\nu + \delta\partial_\mu\phi\partial^\mu\phi + \epsilon\partial_\mu A^\mu\phi \\ &+ \zeta B_\mu B^\mu + \eta\phi^2\Big]\,,\end{aligned} \tag{13}$$

in which A_μ has been set up with the same transverse gauge invariance as in Eq. (2), while B_μ admits the full kinetic structure of a vector field (i.e. no symmetry whatever). In addition to these vectors, a scalar field ϕ is also included in such a way as to preserve the gauge invariance of A_μ. The blocks of the wave operator associated with Eq. (13) are

$$\mathsf{O}_{0^+} = \begin{bmatrix} & A^{\#1}_{0^+} & \phi^{\#1}_{0^+} & B^{\#1}_{0^+} \\ A^{\#1\,\dagger}_{0^+} & k^2\alpha & -\tfrac{1}{2}ik\epsilon & 0 \\ \phi^{\#1\,\dagger}_{0^+} & \tfrac{1}{2}ik\epsilon & k^2\delta+\eta & 0 \\ B^{\#1\,\dagger}_{0^+} & 0 & 0 & (\beta+\gamma)k^2+\zeta \end{bmatrix}, \tag{14a}$$

$$\mathsf{O}_{1^-} = \begin{bmatrix} & A^{\#1\,i}_{1^-} & B^{\#1\,i}_{1^-} \\ A^{\#1\,\dagger}_{1^-\,i} & 0 & 0 \\ B^{\#1\,\dagger}_{1^-\,i} & 0 & \zeta+k^2\gamma \end{bmatrix}. \tag{14b}$$

The 0^+ sector in Eq. (14a) propagates a pair of scalars with square masses $\left(\epsilon^2 - 4\alpha\eta\right)/4\alpha\delta$ and $-\zeta/(\beta+\gamma)$, while the 1^- sector in Eq. (14b) propagates a vector with square mass $-\zeta/\gamma$, and is moreover singular. The residues of the saturated propagator reveal that unitarity for this model is impossible without fine-tuning of the constants, however this is inessential. The MAS may be relied upon to perform the Dirac–Bergmann analysis:

MAS output

This total Hamiltonian describes a quadratic field theory on flat spatial slices with five canonical configuration variables, namely the scalars ϕ, $A^{\#1}_{0^+}$, $B^{\#1}_{0^+}$ and the spatial covectors $A^{\#1}_{1^-i}$, $B^{\#1}_{1^-i}$, together with their conjugate momenta $\pi(\phi)^{\#1}_{0^+}$, $\pi(A)^{\#1}_{0^+}$, $\pi(B)^{\#1}_{0^+}$, $\pi(A)^{\#1\,i}_{1^-}$, and $\pi(B)^{\#1\,i}_{1^-}$, and constant symbols $\alpha, \epsilon, \zeta, \delta, \beta, \eta, \gamma$. The only Lagrange multiplier entering the given Hamiltonian

is $\lambda(A)^{\#1\,i}_{1^-}$, appearing linearly as $\pi(A)^{\#1}_{1^-\,i}\lambda(A)^{\#1\,i}_{1^-}$.
The primary-constraint analysis by variational differentiation with respect to the listed multipliers yields a single nontrivial primary constraint from $\lambda(A)^{\#1}_{1^-\,i}$, namely $\pi(A)^{\#1\,i}_{1^-} \approx 0$, which is a contravariant spatial vector and therefore corresponds to 3 independent constrained degrees of freedom (one per spatial component). All other multiplier variations vanish, i.e. there are no primary constraints associated with $\lambda(\phi)^{\#1}_{0^+}$, $\lambda(A)^{\#1}_{0^+}$, $\lambda(B)^{\#1}_{0^+}$, or $\lambda(B)^{\#1}_{1^-\,i}$ for the Hamiltonian provided.
Preservation in time of the primary constraint gives the condition $\partial^i\pi(A)^{\#1}_{0^+} \approx 0$. Since this is a pure spatial gradient of the scalar $\pi(A)^{\#1}_{0^+}$, the consistency condition is equivalent (modulo spatial zero-modes, as in the standard local Dirac analysis) to the scalar secondary constraint $\pi(A)^{\#1}_{0^+} \approx 0$, corresponding to 1 independent constrained degree of freedom. The secondary constraint is preserved in time without generating further constraints.
All Poisson brackets among the constraints vanish (in particular, the brackets between $\pi(A)^{\#1\,i}_{1^-}$ and $\pi(A)^{\#1}_{0^+}$, as well as each self-bracket, vanish), so both constraints are FC: the 3-component vector constraint $\pi(A)^{\#1\,i}_{1^-} \approx 0$ is FC, and the 1-component scalar constraint $\pi(A)^{\#1}_{0^+} \approx 0$ is FC. No consistency condition fixes the time-evolution of any Lagrange multiplier; in particular, $\lambda(A)^{\#1\,i}_{1^-}$ remains undetermined by the Dirac-Bergmann algorithm for this Hamiltonian, and the remaining listed multipliers (which do not enter the Hamiltonian provided) are likewise undetermined.
The configuration-space dimension is $\frac{1}{2}N_{\rm Can} = 3\,(\text{ from } \phi, A^{\#1}_{0^+}, B^{\#1}_{0^+}) + 6\,(\text{ from } A^{\#1}_{1^-\,i} \text{ and } B^{\#1}_{1^-\,i}) = 9$, so the phase-space dimension is $N_{\rm Can} = 18$. With $N_{\rm FC} = 4$ FC constraint components in total (3 from $\pi(A)^{\#1\,i}_{1^-} \approx 0$ and 1 from $\pi(A)^{\#1}_{0^+} \approx 0$) and $N_{\rm SC} = 0$ SC constraints, the number of physical degrees of freedom is $\frac{1}{2}(N_{\rm Can} - 2N_{\rm FC} - N_{\rm SC}) = \frac{1}{2}(18-8-0) = 5$.

Thus, the MAS correctly reproduces results matching those of Eqs. (14a) and (14b).

Massless multi-gauge theory. — We now consider the theory

$$\begin{aligned}\mathscr{S} = \int \mathrm{d}^4x \Big[&\alpha\partial_\mu A^\mu \partial_\nu A^\nu + \beta\partial_\mu B_\nu \partial^\mu B^\nu \\ &- \beta\partial_\mu B_\nu \partial^\nu B^\mu + \gamma\partial_\mu B_\nu \partial^\mu C^\nu \\ &- \gamma\partial_\mu B_\nu \partial^\nu C^\mu + \delta\partial_\mu C_\nu \partial^\mu C^\nu \\ &- \delta\partial_\mu C_\nu \partial^\nu C^\mu + \delta\partial_\mu\phi\partial^\mu\phi \\ &+ \epsilon\partial_\mu A^\mu\phi + \zeta\phi^2 \Big], \end{aligned} \tag{15}$$

in which the A_μ field has the same role as in Eq. (13), but the B_μ and C_μ fields are now both endowed with the full Maxwell kinetic structure, and are mixed. The scalar field ϕ is again included. The theory in Eq. (15) has a richer gauge structure than that of Eq. (13), spanning multiple fields. The blocks of the wave operator associated with Eq. (15) are

$$\mathsf{O}_{0^+} = \begin{bmatrix} & \phi^{\#1}_{0^+} & A^{\#1}_{0^+} & B^{\#1}_{0^+} & C^{\#1}_{0^+} \\ \phi^{\#1\dagger}_{0^+} & \zeta + k^2\delta & \frac{1}{2}ik\epsilon & 0 & 0 \\ A^{\#1\dagger}_{0^+} & -\frac{1}{2}ik\epsilon & k^2\alpha & 0 & 0 \\ B^{\#1\dagger}_{0^+} & 0 & 0 & 0 & 0 \\ C^{\#1\dagger}_{0^+} & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{16a}$$

$$\mathsf{O}_{1^-} = \begin{bmatrix} & A^{\#1\,i}_{1^-} & B^{\#1\,i}_{1^-} & C^{\#1\,i}_{1^-} \\ A^{\#1\dagger}_{1^-\,i} & 0 & 0 & 0 \\ B^{\#1\dagger}_{1^-\,i} & 0 & k^2\beta & \frac{1}{2}k^2\gamma \\ C^{\#1\dagger}_{1^-\,i} & 0 & \frac{1}{2}k^2\gamma & k^2\delta \end{bmatrix}. \tag{16b}$$

Whilst Eq. (16a) leads to a propagating scalar with square mass $\left(\epsilon^2 - 4\alpha\zeta\right)/4\alpha\delta$, the combined effects of Eqs. (16a) and (16b) in the lightcone limit reveal two pairs of massless polarizations. The two longitudinal (Maxwell-like) gauge symmetries can be read off from Eq. (16a), and the transverse gauge symmetry from Eq. (16b). Unitarity for this model is again impossible without fine-tuning of the constants. Once again, we will rely entirely on the MAS for the actual Hamiltonian structure. The response is:

MAS output

This total Hamiltonian describes a theory on flat spatial slices with seven canonical configuration variables, namely the scalars ϕ, $A^{\#1}_{0^+}$, $B^{\#1}_{0^+}$, $C^{\#1}_{0^+}$ and the spatial covectors $A^{\#1}_{1^-\,i}$, $B^{\#1}_{1^-\,i}$, $C^{\#1}_{1^-\,i}$, together with their conjugate momenta $\pi(\phi)^{\#1}_{0^+}$, $\pi(A)^{\#1}_{0^+}$, $\pi(B)^{\#1}_{0^+}$, $\pi(C)^{\#1}_{0^+}$ and $\pi(A)^{\#1\,i}_{1^-}$, $\pi(B)^{\#1\,i}_{1^-}$, $\pi(C)^{\#1\,i}_{1^-}$. The constant symbols entering the Hamiltonian are α, ϵ, ζ, δ, β, and γ.
The Dirac–Bergmann analysis yields three primary constraints. The first is the vector constraint $\pi(A)^{\#1\,i}_{1^-} \approx 0$, which has 3 independent components. The second is the scalar constraint $\pi(B)^{\#1}_{0^+} \approx 0$, which has 1 independent component. The third is the scalar constraint $\pi(C)^{\#1}_{0^+} \approx 0$, which has 1 independent component. Preservation in time of $\pi(A)^{\#1\,i}_{1^-} \approx 0$ yields a secondary constraint via a spatial gradient, which reduces to the scalar constraint $\pi(A)^{\#1}_{0^+} \approx 0$ with 1 independent component. Preservation in time of $\pi(B)^{\#1}_{0^+} \approx 0$ yields the secondary scalar constraint $\partial_i\pi(B)^{\#1\,i}_{1^-} \approx 0$ with 1 independent component. Preservation in time of $\pi(C)^{\#1}_{0^+} \approx 0$ yields the secondary scalar constraint $\partial_i\pi(C)^{\#1\,i}_{1^-} \approx 0$ with 1 independent component. No tertiary constraints arise: the time evolution of $\pi(A)^{\#1}_{0^+}$, $\partial_i\pi(B)^{\#1\,i}_{1^-}$, and $\partial_i\pi(C)^{\#1\,i}_{1^-}$ vanishes on the constraint surface.
All six constraints listed above are FC: all computed mutual Poisson brackets between smeared forms of these constraints vanish. No SC constraints occur. Consequently, no Lagrange multiplier is fixed by consistency conditions. In particular, the multipliers that explicitly enforce primary constraints in the given to-

tal Hamiltonian, namely $\lambda(A)^{\#1}_{1^-i}$, $\lambda(B)^{\#1}_{0^+}$, and $\lambda(C)^{\#1}_{0^+}$, remain undetermined by the Dirac–Bergmann algorithm. The remaining multiplier symbols provided in the model input, $\lambda(\phi)^{\#1}_{0^+}$, $\lambda(A)^{\#1}_{0^+}$, $\lambda(B)^{\#1}_{1^-i}$, and $\lambda(C)^{\#1}_{1^-i}$, do not generate primary constraints for this Hamiltonian and are likewise not determined by consistency conditions.
The physical degree-of-freedom count is as follows. The configuration space has 13 independent components (4 scalars plus 3 vectors, i.e. $4+3+3+3=13$), so the phase space has dimension $N_{\rm Can}=26$. There are $N_{\rm FC}=8$ independent FC constraints in total (3 from $\pi(A)^{\#1\,i}_{1^-}\approx 0$, and 1 each from $\pi(B)^{\#1}_{0^+}\approx 0$, $\pi(C)^{\#1}_{0^+}\approx 0$, $\pi(A)^{\#1}_{0^+}\approx 0$, $\partial_i\pi(B)^{\#1\,i}_{1^-}\approx 0$, and $\partial_i\pi(C)^{\#1\,i}_{1^-}\approx 0$), and $N_{\rm SC}=0$ SC constraints. Therefore the number of physical degrees of freedom is $(N_{\rm Can}-2N_{\rm FC}-N_{\rm SC})/2=(26-16-0)/2=5$.

Thus, the MAS correctly navigates multiple constraint chains to reproduce results matching those of Eqs. (16a) and (16b).

Li–Gao theory. — We now consider a model which was put forwards very recently in [73], namely

$$\begin{aligned}\mathscr{S}=\int \mathrm{d}^4x\,\Big[&(\beta_3+\gamma_3)\,\dot{A}^{\#1}_{1^-i}\partial^i A^{\#1}_{0^+}+\alpha_3\dot{A}^{\#1}_{1^-i}\dot{A}^{\#1\,i}_{1^-}\\&+\frac{(\beta_3+\gamma_3)^2-\beta_3\gamma_3}{4\alpha_3}\partial_i A^{\#1}_{0^+}\partial^i A^{\#1}_{0^+}\\&+(c_{4,3}+c_{4,5})\,\partial_i A^{\#1}_{1^-j}\partial^j A^{\#1\,i}_{1^-}\\&+c_{4,4}\partial_i A^{\#1}_{1^-j}\partial^i A^{\#1\,j}_{1^-}\Big]\,.\end{aligned}\tag{17}$$

The theory in Eq. (17) only depends on the field A_μ, but it is deliberately inexpressible in Lorentz-covariant form. In such cases, spin-projection is no longer the natural approach and we rely on the Hamiltonian analysis in [73] to determine the spectrum.[7] According to the MAS:

MAS output

This total Hamiltonian describes a theory on flat spatial slices with canonical configuration variables $A^{\#1}_{0^+}$ (one component) and $A^{\#1}_{1^-i}$ (three components), with conjugate momenta $\pi(A)^{\#1}_{0^+}$ and $\pi(A)^{\#1\,i}_{1^-}$, and constant symbols $\alpha_3,\beta_3,c_{4,3},c_{4,4},c_{4,5},\gamma_3$. Varying the total Hamiltonian density with respect to the multipliers shows that $\lambda(A)^{\#1}_{0^+}$ enforces the (scalar) primary constraint $\pi(A)^{\#1}_{0^+}\approx 0$, which contains exactly one independent constrained degree of freedom. Variation with respect to $\lambda(A)^{\#1}_{1^-i}$ yields identically zero, so there is no primary constraint associated with that multiplier. Preservation in time of the primary constraint yields a nontrivial secondary condition which does not contain any Lagrange multiplier, and hence is a genuine secondary constraint. In canonical variables it is the scalar equation $-\frac{1}{2}{\alpha_3}^{-1}(\beta_3+\gamma_3)\partial_i\pi(A)^{\#1\,i}_{1^-}-\frac{1}{2}{\alpha_3}^{-1}\beta_3\gamma_3\partial_i\partial^i A^{\#1}_{0^+}\approx 0$, which again contains exactly one independent constrained degree of freedom. Preservation in time of this secondary constraint holds identically (its Poisson bracket with the total Hamiltonian vanishes), so the Dirac–Bergmann algorithm terminates with precisely these two constraints.
The mutual Poisson bracket between the smeared primary and smeared secondary constraints vanishes, and with no further constraints present this implies that both constraints are FC. Since the consistency conditions generate no multiplier equations, no Lagrange multiplier is determined by time preservation; the remaining undetermined multipliers are $\lambda(A)^{\#1}_{0^+}$ (which multiplies the primary constraint in the total Hamiltonian) and $\lambda(A)^{\#1}_{1^-i}$ (which remains arbitrary and does not enter the given total Hamiltonian density).
The total number of canonical degrees of freedom is $2\,(1+3)=8$. With two FC constraints and no SC constraints, the physical degrees of freedom are $(8-2\times 2-0)/2=2$.

The MAS confirms the findings of [73], namely that Eq. (17) shares its spectrum with the Maxwell theory. This feature, despite the global Lorentz symmetry being broken, is put forwards in [73] as having some utility in cosmological model-building.

III CONCLUSIONS

Overview. — It was demonstrated that a simple MAS is capable of correctly performing the Dirac–Bergmann algorithm from start to finish, without user intervention. The system shown here is restricted to free theories on Minkowski spacetime, but is agnostic regarding field content and gauge structure. These results are significant, because the Dirac–Bergmann algorithm is a staple tool in the theoretical high-energy physics and relativity communities. It has, moreover, resisted automation until now, owing to its dual reliance on low-level symbolic manipulation and high-level reasoning. The symbolic aspect was recently addressed with conventional computer algebra techniques [53]. In this work the reasoning aspect was shown to be within the scope of commercially available LLMs.

Dirac–Bergmann is not 'special'. — Despite having a somewhat formidable reputation in the relativity community, we find no signal that the Dirac–Bergmann algorithm is inherently challenging as a natural language problem. Indeed, the architecture in Fig. 1, which has been shown to suffice for free theories, may be termed 'embarrassingly simple'. For example:

- While meeting the criteria for a MAS [74], the imple-

[7] Note that [73] was announced two months after the training cutoff of **gpt-5.2**.

mentation may equally be viewed as a single-agent system plus a control loop. Performance may be improved by with multiple, more highly specialised LLM agents, possibly at the expense of an inference penalty.

- The implementation relies entirely on in-context learning by a closed model. Performance may be improved by domain-specific fine-tuning of open-weight models, to be balanced against a training penalty.

There is thus ample room for exploring more sophisticated agentic designs in response to rising demands. Since the computational tools in [53] are already applicable to interacting and non-linear gravity theories such as in [43, 59, 60], it seems plausible that a successful MAS may indeed be developed for these domains.

Test-time scaling. — It is critical to remember that our results are specific to **`gpt-5.2`**, and the architecture in Fig. 1 is strongly influenced by the test-time scaling features inherent to that model. As *OpenAI*'s flagship LLM, these features are presumably advanced, but they are not openly documented: the **`reasoning_effort`** parameter significantly affects performance and cost, but offers little direct control. As a result, both the REPL agent and system prompt are largely devoted to 'buffering' common failure modes with natural language so as to trigger adaptive reasoning. Our reliance on closed commercial models offers high performance at low cost; enough to demonstrate the viability of the approach as a proof of concept (which is our main goal). It does not, however, offer insight into the Dirac–Bergmann algorithm as a test-time scaling problem. It is this latter question which must be addressed in order to understand how LLMs may be deployed for scientific discovery in theoretical physics [68].

Symbolic verification and tool usage. — Control over the validity of results is essential. For applications to free theories, we have demonstrated how performance can be tested against a curated catalogue of models, whose spectra are determined independently via the spin-projection method (or from prior literature [73]). This approach does *not* scale to novel interacting and non-linear gravity theories, or to scientific discovery in general. Nonetheless, the reliance of the system on conventional computer algebra tools [53–57] greatly facilitates validation, and this principle *does* scale. Tool-calls are logged, allowing a minimal walk-through of the kernel session to be extracted. From the user perspective, the cognitive burden of assessing the validity of a start-to-finish run of the Dirac–Bergmann algorithm is generally much lower than that of performing the algorithm ab initio. This is especially true given that the cost of computing and simplifying the Poisson brackets is removed. As an alternative to user verification, there is scope for the use of weak verification tools. Prominent failure modes included the inability to identify (i) dependent constraints, (ii) FC combinations, and (iii) fewer constrained degrees of freedom than naïvely suggested by the index structure. Extension to interactions and non-linear gravity will likely exacerbate these issues. Once the LLM agent makes a type-(i) or type-(ii) identification, the tools in [53–57] (which employ a state-of-the-art implementation of the Butler–Portugal algorithm [75–77][8]) will almost never fail to verify or refute it. These tools do not help with type-(iii) identifications, or with the *failure* to make identifications, both of which reduce to multi-term tensor algebra problems for which there is no algorithm. In these cases, however, a promising avenue to weak verification is the use of numerical sampling, as employed for the Dirac–Bergmann algorithm in [81]. This weak verification approach will be explored in future work.

Note of caution. — The MAS presented in this work does not have the functionality of an interactive assistant. The goal is to determine whether the Dirac–Bergmann algorithm may be automated in its entirety, without requiring user intervention. This end-to-end problem is interesting precisely because it is seems hard, whilst also being useful to specific theory communities. We do not put forward this MAS or future systems as 'replacements' for the user. A viable Dirac–Bergmann system for interacting and non-linear gravity theories would be a powerful tool for augmenting the existing capabilities of high-energy physics and relativity researchers. It would not, however, obviate the need for human expertise.

ACKNOWLEDGEMENTS

This work was made possible by useful discussions with Boris Bolliet, Justin Feng, Will Handley, Tobias Mistele, Sebastian Murk, Syksy Räsänen, Ignacy Sawicki and Tom Zlosnik.

WB is grateful for the support of Marie Skłodowska-Curie Actions and the Institute of Physics of the Czech Academy of Sciences.

WB was supported by the research environment and infrastructure of the Handley Lab at the University of Cambridge.

This work was performed using the Cambridge Service for Data Driven Discovery (CSD3), part of which is operated by the University of Cambridge Research Computing on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). The DiRAC component of CSD3 was funded by BEIS capital funding via STFC capital grants ST/P002307/1 and ST/R002452/1 and STFC operations grant ST/R00689X/1. DiRAC is part of the National e-Infrastructure.

Co-funded by the European Union (Physics for Future – Grant Agreement No. 101081515). Views and opinions

[8] This is the reason why we rely on the proprietary *Wolfram* system, rather than directly on *SymPy*, although it is to be noted that the open-source *Cadabra* system [78–80] uses the same canonicalizer as *xAct* [77].

expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

[1] S. Deser, Gen. Rel. Grav. **1**, 9 (1970), arXiv:gr-qc/0411023.
[2] S. Deser, Class. Quant. Grav. **4**, L99 (1987).
[3] T. Padmanabhan, Int. J. Mod. Phys. D **17**, 367 (2008), arXiv:gr-qc/0409089.
[4] L. M. Butcher, M. Hobson, and A. Lasenby, Phys. Rev. D **80**, 084014 (2009), arXiv:0906.0926 [gr-qc].
[5] R. Utiyama, Phys. Rev. **101**, 1597 (1956).
[6] A. K. H. Bengtsson, Phys. Lett. B **182**, 321 (1986).
[7] M. Henneaux and C. Teitelboim, in *2nd Meeting on Quantum Mechanics of Fundamental Systems (CECS)* (1987).
[8] D. Francia and A. Sagnotti, Class. Quant. Grav. **20**, S473 (2003), arXiv:hep-th/0212185.
[9] A. Sagnotti and M. Tsulaia, Nucl. Phys. B **682**, 83 (2004), arXiv:hep-th/0311257.
[10] A. Fotopoulos and M. Tsulaia, Int. J. Mod. Phys. A **24**, 1 (2009), arXiv:0805.1346 [hep-th].
[11] B. Holdom, Phys. Lett. B **166**, 196 (1986).
[12] R. D. Peccei and H. R. Quinn, Phys. Rev. Lett. **38**, 1440 (1977).
[13] S. Weinberg, Phys. Rev. Lett. **40**, 223 (1978).
[14] F. Wilczek, Phys. Rev. Lett. **40**, 279 (1978).
[15] C. Csáki, S. Lombardo, and O. Telem, in *Theoretical Advanced Study Institute in Elementary Particle Physics: Anticipating the Next Discoveries in Particle Physics* (WSP, 2018) pp. 501–570, arXiv:1811.04279 [hep-ph].
[16] G. Isidori, F. Wilsch, and D. Wyler, Rev. Mod. Phys. **96**, 015006 (2024), arXiv:2303.16922 [hep-ph].
[17] P. Van Nieuwenhuizen, Nucl. Phys. B **60**, 478 (1973).
[18] D. E. Neville, Phys. Rev. D **18**, 3535 (1978).
[19] D. E. Neville, Phys. Rev. D **21**, 867 (1980).
[20] E. Sezgin, Phys. Rev. D **24**, 1677 (1981).
[21] E. Sezgin and P. van Nieuwenhuizen, Phys. Rev. D **21**, 3269 (1980).
[22] R. Kuhfuss and J. Nitsch, Gen. Rel. Grav. **18**, 1207 (1986).
[23] G. K. Karananas, Class. Quant. Grav. **32**, 055012 (2015), arXiv:1411.5613 [gr-qc].
[24] G. K. Karananas, *Poincaré, Scale and Conformal Symmetries Gauge Perspective and Cosmological Ramifications*, Ph.D. thesis, Ecole Polytechnique, Lausanne (2016), arXiv:1608.08451 [hep-th].
[25] E. L. Mendonça and R. Schimidt Bittencourt, Adv. High Energy Phys. **2020**, 8425745 (2020), arXiv:1902.05118 [hep-th].
[26] R. Percacci and E. Sezgin, Phys. Rev. D **101**, 084040 (2020), arXiv:1912.01023 [hep-th].
[27] R. Percacci and E. Sezgin, Phys. Rev. D **101**, 084040 (2020), arXiv:1912.01023 [hep-th].
[28] C. Marzo, Phys. Rev. D **105**, 065017 (2022), arXiv:2108.11982 [hep-ph].
[29] C. Marzo, Phys. Rev. D **106**, 024045 (2022), arXiv:2110.14788 [hep-th].
[30] Y. Mikura, V. Naso, and R. Percacci, Phys. Rev. D **109**, 104071 (2024), arXiv:2312.10249 [gr-qc].
[31] Y. Mikura and R. Percacci, (2024), arXiv:2401.10097 [gr-qc].
[32] A. Aurilia and H. Umezawa, Phys. Rev. **182**, 1682 (1969).
[33] L. Buoninfante, (2016), arXiv:1610.08744 [gr-qc].
[34] Y.-C. Lin, M. P. Hobson, and A. N. Lasenby, Phys. Rev. D **99**, 064001 (2019), arXiv:1812.02675 [gr-qc].
[35] Y.-C. Lin, M. P. Hobson, and A. N. Lasenby, Phys. Rev. D **101**, 064038 (2020), arXiv:1910.14197 [gr-qc].
[36] Y.-C. Lin, M. P. Hobson, and A. N. Lasenby, Phys. Rev. D **104**, 024034 (2021), arXiv:2005.02228 [gr-qc].
[37] Y.-C. Lin, *Ghost and tachyon free gauge theories of gravity: A systematic approach*, Ph.D. thesis, Cambridge U. (2020).
[38] W. Barker, C. Marzo, and C. Rigouzzo, Phys. Rev. D **112**, 016018 (2025), arXiv:2406.09500 [hep-th].
[39] W. Barker, G. K. Karananas, and H. Tu, Phys. Rev. D **112**, 084041 (2025), arXiv:2506.02111 [hep-th].
[40] D. Hutchings and M. Ponds, JHEP **07**, 292, arXiv:2401.04523 [hep-th].
[41] H.-j. Yo and J. M. Nester, Int. J. Mod. Phys. D **8**, 459 (1999), arXiv:gr-qc/9902032.
[42] H.-J. Yo and J. M. Nester, Int. J. Mod. Phys. D **11**, 747 (2002), arXiv:gr-qc/0112030.
[43] W. Barker and D. Glavan, (2025), arXiv:2510.08201 [gr-qc].
[44] P. A. M. Dirac, Can. J. Math. **2**, 129 (1950).
[45] P. A. M. Dirac, Proc. Roy. Soc. Lond. A **246**, 326 (1958).
[46] P. G. Bergmann and I. Goldberg, Phys. Rev. **98**, 531 (1955).
[47] J. L. Anderson and P. G. Bergmann, Phys. Rev. **83**, 1018 (1951).
[48] L. Castellani, Annals Phys. **143**, 357 (1982).
[49] M. Henneaux and C. Teitelboim, *Quantization of gauge systems* (Princeton University Press, Princeton, New Jersey, 1992).
[50] M. Blagojević, *Gravitation and Gauge Symmetries*, Series in high energy physics, cosmology, and gravitation (Institute of Physics Publishing, Bristol, UK, 2002).
[51] C. Becchi, A. Rouet, and R. Stora, Annals Phys. **98**, 287 (1976).
[52] I. V. Tyutin, (1975), arXiv:0812.0580 [hep-th].
[53] W. Barker, (2025), arXiv:2512.25007 [physics.comp-ph].
[54] J. M. Martin-Garcia, R. Portugal, and L. R. U. Manssur, Comput. Phys. Commun. **177**, 640 (2007), arXiv:0704.1756 [cs.SC].
[55] J. M. Martín-García, Comput. Phys. Commun. **179**, 597 (2008), arXiv:0803.0862 [cs.SC].
[56] J. M. Martin-Garcia, D. Yllanes, and R. Portugal, Comput. Phys. Commun. **179**, 586 (2008), arXiv:0802.1274 [cs.SC].
[57] T. Nutma, Comput. Phys. Commun. **185**, 1719 (2014), arXiv:1308.3493 [cs.SC].
[58] W. E. V. Barker, Eur. Phys. J. C **83**, 228 (2023), arXiv:2206.00658 [gr-qc].
[59] I. L. Buchbinder and S. L. Lyakhovich, Class. Quant. Grav. **4**, 1487 (1987).
[60] D. Glavan, S. Mukohyama, and T. Zlosnik, JCAP **01**, 111, arXiv:2409.15989 [gr-qc].

[61] J. Kaplan, S. McCandlish, T. Henighan, T. B. Brown, B. Chess, R. Child, S. Gray, A. Radford, J. Wu, and D. Amodei, arXiv e-prints , arXiv:2001.08361 (2020), arXiv:2001.08361 [cs.LG].

[62] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, arXiv e-prints , arXiv:1706.03762 (2017), arXiv:1706.03762 [cs.CL].

[63] A. Laverick, K. Surrao, I. Zubeldia, B. Bolliet, M. Cranmer, A. Lewis, B. Sherwin, and J. Lesgourgues, (2024), arXiv:2412.00431 [astro-ph.IM].

[64] L. Xu, M. Sarkar, A. I. Lonappan, Íñigo Zubeldia, P. Villanueva-Domingo, S. Casas, C. Fidler, C. Amancharla, U. Tiwari, A. Bayer, C. A. Ekioui, M. Cranmer, A. Dimitrov, J. Fergusson, K. Gandhi, S. Krippendorf, A. Laverick, J. Lesgourgues, A. Lewis, T. Meier, B. Sherwin, K. Surrao, F. Villaescusa-Navarro, C. Wang, X. Xu, and B. Bolliet, Open source planning and control system with language agents for autonomous scientific discovery (2025), arXiv:2507.07257 [cs.AI].

[65] S. Casas, C. Fidler, B. Bolliet, F. Villaescusa-Navarro, and J. Lesgourgues, (2025), arXiv:2508.05728 [astro-ph.IM].

[66] F. Villaescusa-Navarro, B. Bolliet, P. Villanueva-Domingo, A. E. Bayer, A. Acquah, C. Amancharla, A. Barzilay-Siegal, P. Bermejo, C. Bilodeau, P. Cárdenas Ramírez, M. Cranmer, U. L. França, C. Hahn, Y.-F. Jiang, R. Jimenez, J.-Y. Lee, A. Lerario, O. Mamun, T. Meier, A. A. Ojha, P. Protopapas, S. Roy, D. N. Spergel, P. Tarancón-Álvarez, U. Tiwari, M. Viel, D. Wadekar, C. Wang, B. Y. Wang, L. Xu, Y. Yovel, S. Yue, W.-H. Zhou, Q. Zhu, J. Zou, and Í. Zubeldia, arXiv e-prints , arXiv:2510.26887 (2025), arXiv:2510.26887 [cs.AI].

[67] D. J. H. Chung, Z. Gao, Y. Kvasiuk, T. Li, M. Münchmeyer, M. Rudolph, F. Sala, and S. C. Tadepalli, Mach. Learn. Sci. Tech. **6**, 030505 (2025), arXiv:2502.15815 [cs.LG].

[68] Z. Gao, T. Li, Y. Kvasiuk, S. C. Tadepalli, M. Rudolph, D. J. H. Chung, F. Sala, and M. Münchmeyer, (2025), arXiv:2506.20729 [cs.LG].

[69] OpenAI, GPT-5.2 Large Language Model, `https://platform.openai.com/docs/guides/latest-model` (2025), latest OpenAI documentation for GPT-5.2, accessed December 2025.

[70] OpenAI, GPT-5 System Card, `https://cdn.openai.com/gpt-5-system-card.pdf` (2025), accessed August 2025.

[71] W. Barker, *Supplemental materials at the GitHub repository www.github.com/wevbarker/Hasdrubal.*

[72] K. Lo, L. L. Wang, M. Neumann, R. Kinney, and D. S. Weld, arXiv e-prints , arXiv:1911.02782 (2019), arXiv:1911.02782 [cs.CL].

[73] S.-Y. Li and X. Gao, (2025), arXiv:2510.26639 [hep-th].

[74] M. Wooldridge, *An Introduction to MultiAgent Systems* (Wiley, 2009).

[75] G. Butler, ed., *Fundamental Algorithms for Permutation Groups*, 1st ed., Lecture Notes in Computer Science, Vol. 559 (Springer, Berlin, Heidelberg, 1991).

[76] L. R. U. Manssur, R. Portugal, and B. F. Svaiter, International Journal of Modern Physics C **13**, 859 (2002), arXiv:math-ph/0107032 [math-ph].

[77] B. E. Niehoff, Computer Physics Communications **228**, 123 (2018).

[78] K. Peeters, Introducing Cadabra: A Symbolic computer algebra system for field theory problems (2007), arXiv:hep-th/0701238 [hep-th].

[79] K. Peeters, Comput. Phys. Commun. **176**, 550 (2007), arXiv:cs/0608005 [cs.SC].

[80] K. Peeters, J. Open Source Softw. **3**, 1118 (2018).

[81] M. Banados and M. Henneaux, (2025), arXiv:2511.20903 [hep-th].