DarkMatterFM: An Agentic Foundation Model for Multimodal Dark-Matter Inference with GPU-Accelerated Emulators

Ioana A. Zelko [1] **Jo Bovy** [1]

[1] *David A. Dunlap Department of Astronomy and Astrophysics, University of Toronto, Toronto, ON, Canada*. Correspondence to: Ioana A. Zelko ioana.zelko@gmail.com.

1. Introduction

Determining the microphysical nature of dark matter (DM) remains a central problem in astrophysics [1]. Key DM scenarios (e.g., warm DM, sterile-neutrino DM, self-interacting DM, ultralight/fuzzy DM) imprint distinct signatures on structure formation and galaxy evolution across a wide range of scales [2, 3, 4, 5]. However, progress is limited by *workflow fragmentation*: each DM model class couples differently to each observational probe (CMB anisotropies, Lyman-α forest flux power, stellar-stream gaps, satellite counts, strong-lensing anomalies), and combining probes typically requires bespoke pipelines and expensive forward modeling [6, 7, 8, 9, 10, 11, 12].

We present **DarkMatterFM**, an *agentic foundation-model workflow* that turns a user-specified DM hypothesis into routed predictions and likelihood evaluations across multiple probes, returning calibrated posteriors and provenance. Conceptually, DarkMatterFM adapts the "foundation model" paradigm [13] to scientific inference, pairing an LLM planner with GPU-accelerated emulators so that multimodal DM constraints become interactive rather than weeks-long.

2. System overview and technical approach

DarkMatterFM is organized as a set of microservices with three coupled layers:

2.1 Theory registry and literature-grounded planning

A curated domain corpus supports retrieval and tool routing. An LLM planner interprets a request into a structured specification: DM model class/parameters, priors and nuisance parameters, applicable probes, and required emulator/likelihood calls. This design follows the broader trend of ML-enabled cosmology workflows [14] while emphasizing *actionable* outputs (schemas, parameter maps, citations) rather than narrative summaries.

2.2 Schema-defined probe data lake

Probe inputs are standardized into versioned, schema-validated datasets including covariances and selection effects where applicable. This includes representative open-source likelihood products for CMB and Lyman-α [6, 7], plus curated summaries for stellar streams and lensing [8, 15, 16, 10, 11], and satellite constraints [9].

2.3 GPU-accelerated emulator + inference stack

We train uncertainty-calibrated emulators to replace costly forward models, motivated by prior successes in cosmological emulation and simulation surrogates [17, 18, 19, 20]. When appropriate, we incorporate physics-informed training signals [21]. Given emulator outputs $\hat{d}(\theta)$ for parameters θ, we evaluate probe likelihoods and perform joint inference:

$$p(\theta \mid D) \propto p(\theta) \prod_{i=1}^{N_{\text{probes}}} \mathcal{L}_i\left(D_i \,\middle|\, \hat{d}_i(\theta)\right). \tag{1}$$

3. Scope: probes, emulators, and targets

Table 1 summarizes the initial probe set and intended emulator interfaces. The goal is not to replace high-fidelity pipelines, but to provide *validated, fast* approximations that enable rapid exploration and principled combination of complementary constraints already demonstrated across probes [12, 10, 8, 9].

Table 1: Initial DarkMatterFM probe interfaces and typical latency targets.

Probe	Observable	Emulator output	Target latency
CMB	spectra/likelihood	$\hat{C}_\ell(\theta)$ or summaries	$\lesssim$ seconds
Lyα	flux power	$\widehat{P}_F(k, z; \theta)$	$\lesssim$ seconds
Streams	gap statistics	rate/shape summaries	$\lesssim$ seconds
Satellites	counts	HMF/SHMF summaries	$\lesssim$ seconds
Quads	lensing anomalies	substructure summaries	$\lesssim$ seconds

4. Validation plan and expected outcomes

DarkMatterFM is evaluated along three axes:

Emulator fidelity and calibration. We benchmark emulators against held-out forward runs and report accuracy metrics relevant to inference (errors in summary statistics, coverage tests, simulation-based calibration), following best practices in cosmological surrogate modeling [17, 18, 19].

End-to-end inference correctness. For representative DM models (e.g., sterile-neutrino, warm DM, fuzzy DM, SIDM), we compare DarkMatterFM posteriors against established single-probe analyses and multi-probe combinations [10, 8, 12, 3, 4, 2].

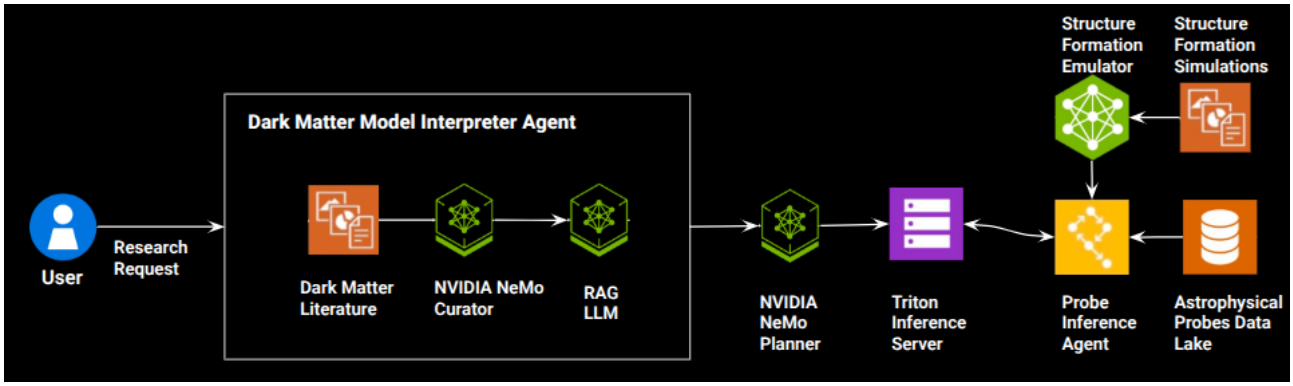

Fig. 1: **DarkMatterFM architecture.** DarkMatterFM concept: an LLM planner interprets a DM hypothesis, routes to probe-specific emulators and likelihoods served as microservices, and returns joint constraints with provenance. A user query is parsed by an LLM planner (literature-grounded RAG) into a structured DM model specification and a set of relevant probes. The planner routes execution to Triton-served microservices: (i) GPU emulators for structure formation / probe summaries and (ii) probe likelihood evaluators (CMB, Lyα, streams, satellites, quads). Outputs are combined into joint posteriors and Bayes factors with full provenance (data versions, emulator checkpoints, and tool traces) to enable auditable, reproducible multimodal inference.

Systems performance and reproducibility. We measure serving latency/throughput and ensure auditable provenance (versioned data, emulator checkpoints, and planner traces), enabling reuse by the wider community.

The AI4X deliverable is a working, extensible system blueprint with proof-of-concept multimodal constraints, demonstrating how agentic planning + GPU emulation can convert a traditionally weeks-long analysis loop into an interactive workflow for dark-matter physics.

4.1 Related work

DarkMatterFM sits at the intersection of (i) foundation-model framing for broad task coverage [13], (ii) physics-informed learning [21], and (iii) cosmological emulation/surrogates [20, 19, 18, 17]. Its novelty is the explicit *theory-to-probe routing* layer and the "served" multimodal inference mindset, aimed at making joint DM constraints routine rather than exceptional [12].

Acknowledgments

This work is supported by the University of Toronto's Eric and Wendy Schmidt AI in Science Postdoctoral Fellowship, a program of Schmidt Sciences (I. A. Z.) and by the Natural Sciences & Engineering Research Council of Canada (NSERC), (funding reference number RGPIN-2020-04712, J. B.). We thank the community tool developers and open-data efforts that enable multimodal cosmological inference.

References

[1] Gianfranco Bertone and Dan Hooper. History of dark matter. 90(4):045002. _eprint: 1605.04909.

[2] A. Boyarsky, M. Drewes, T. Lasserre, S. Mertens, and O. Ruchayskiy. Sterile neutrino dark matter. 104:1–45. _eprint: 1807.07938.

[3] Sean Tulin and Hai-Bo Yu. Dark matter self-interactions and small scale structure. 730:1–57. _eprint: 1705.02358.

[4] Lam Hui, Jeremiah P. Ostriker, Scott Tremaine, and Edward Witten. Ultralight scalars as cosmological dark matter. 95(4):043541. _eprint: 1610.08297.

[5] James S. Bullock and Michael Boylan-Kolchin. Small-scale challenges to the \ensuremathŁambdaCDM paradigm. 55(1):343–387. _eprint: 1707.04256.

[6] Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, S. Basak, R. Battye, K. Benabed, J. P. Bernard, M. Bersanelli, P. Bielewicz, J. J. Bock, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, J. F. Cardoso, J. Carron, A. Challinor, H. C. Chiang, J. Chluba, L. P. L. Colombo, C. Combet, D. Contreras, B. P. Crill, F. Cuttaia, P. de Bernardis, G. de Zotti, J. Delabrouille, J. M. Delouis, E. Di Valentino, J. M. Diego, O. Doré, M. Douspis, A. Ducout, X. Dupac, S. Dusini, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, Y. Fantaye, M. Farhang, J. Fergusson, R. Fernandez-Cobos, F. Finelli, F. Forastieri, M. Frailis, A. A. Fraisse, E. Franceschi, A. Frolov, S. Galeotta, S. Galli, K. Ganga, R. T. Génova-Santos, M. Gerbino, T. Ghosh, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gruppuso, J. E. Gudmundsson, J. Hamann, W. Handley, F. K. Hansen, D. Herranz, S. R. Hildebrandt, E. Hivon, Z. Huang, A. H. Jaffe, W. C. Jones, A. Karakci, E. Keihänen, R. Keskitalo, K. Kiiveri, J. Kim, T. S. Kisner, L. Knox, N. Krachmalnicoff, M. Kunz, H. Kurki-Suonio, G. Lagache, J. M. Lamarre, A. Lasenby, M. Lattanzi, C. R.

Lawrence, M. Le Jeune, P. Lemos, J. Lesgourgues, F. Levrier, A. Lewis, M. Liguori, P. B. Lilje, M. Lilley, V. Lindholm, M. López-Caniego, P. M. Lubin, Y. Z. Ma, J. F. Macías-Pérez, G. Maggio, D. Maino, N. Mandolesi, A. Mangilli, A. Marcos-Caballero, M. Maris, P. G. Martin, M. Martinelli, E. Martínez-González, S. Matarrese, N. Mauri, J. D. McEwen, P. R. Meinhold, A. Melchiorri, A. Mennella, M. Migliaccio, M. Millea, S. Mitra, M. A. Miville-Deschênes, D. Molinari, L. Montier, G. Morgante, A. Moss, P. Natoli, H. U. Nørgaard-Nielsen, L. Pagano, D. Paoletti, B. Partridge, G. Patanchon, H. V. Peiris, F. Perrotta, V. Pettorino, F. Piacentini, L. Polastri, G. Polenta, J. L. Puget, J. P. Rachen, M. Reinecke, M. Remazeilles, A. Renzi, G. Rocha, C. Rosset, G. Roudier, J. A. Rubiño-Martín, B. Ruiz-Granados, L. Salvati, M. Sandri, M. Savelainen, D. Scott, E. P. S. Shellard, C. Sirignano, G. Sirri, L. D. Spencer, R. Sunyaev, A. S. Suur-Uski, J. A. Tauber, D. Tavagnacco, M. Tenti, L. Toffolatti, M. Tomasi, T. Trombetti, L. Valenziano, J. Valiviita, B. Van Tent, L. Vibert, P. Vielva, F. Villa, N. Vittorio, B. D. Wandelt, I. K. Wehus, M. White, S. D. M. White, A. Zacchei, and A. Zonca. Planck 2018 results. VI. cosmological parameters. 641:A6. _eprint: 1807.06209.

[7] Vid Iršič, Matteo Viel, Martin G. Haehnelt, James S. Bolton, Stefano Cristiani, George D. Becker, Valentina D'Odorico, Guido Cupani, Tae-Sun Kim, Trystyn A. M. Berg, Sebastian López, Sara Ellison, Lise Christensen, Kelly D. Denney, and Gábor Worseck. New constraints on the free-streaming of warm dark matter from intermediate and small scale lyman-\ensuremath\alpha forest data. 96(2):023522. _eprint: 1702.01764.

[8] Nilanjan Banik, Jo Bovy, Gianfranco Bertone, Denis Erkal, and T. J. L. de Boer. Novel constraints on the particle nature of dark matter from stellar streams. 2021:043. Publisher: IOP ADS Bibcode: 2021JCAP...10..043B.

[9] E. O. Nadler, A. Drlica-Wagner, K. Bechtol, S. Mau, R. H. Wechsler, V. Gluscevic, K. Boddy, A. B. Pace, T. S. Li, M. McNanna, A. H. Riley, J. García-Bellido, Y. Y. Mao, G. Green, D. L. Burke, A. Peter, B. Jain, T. M. C. Abbott, M. Aguena, S. Allam, J. Annis, S. Avila, D. Brooks, M. Carrasco Kind, J. Carretero, M. Costanzi, L. N. da Costa, J. De Vicente, S. Desai, H. T. Diehl, P. Doel, S. Everett, A. E. Evrard, B. Flaugher, J. Frieman, D. W. Gerdes, D. Gruen, R. A. Gruendl, J. Gschwend, G. Gutierrez, S. R. Hinton, K. Honscheid, D. Huterer, D. J. James, E. Krause, K. Kuehn, N. Kuropatkin, O. Lahav, M. A. G. Maia, J. L. Marshall, F. Menanteau, R. Miquel, A. Palmese, F. Paz-Chinchón, A. A. Plazas, A. K. Romer, E. Sanchez, V. Scarpine, S. Serrano, I. Sevilla-Noarbe, M. Smith, M. Soares-Santos, E. Suchyta, M. E. C. Swanson, G. Tarle, D. L. Tucker, A. R. Walker, W. Wester, and DES Collaboration. Constraints on dark matter properties from observations of milky way satellite galaxies. 126(9):091101. _eprint: 2008.00022.

[10] Ioana A. Zelko, Tommaso Treu, Kevork N. Abazajian, Daniel Gilman, Andrew J. Benson, Simon Birrer, Anna M. Nierenberg, and Alexander Kusenko. Constraints on sterile neutrino models from strong gravitational lensing, milky way satellites, and the lyman-α forest. 129:191301. Publisher: APS ADS Bibcode: 2022PhRvL.129s1301Z.

[11] Ioana A. Zelko, Anna M. Nierenberg, and Tommaso Treu. Probing dark matter with adaptive-optics based flux ratio anomalies: photometric and astrometric precision. 531:885–897. Publisher: OUP ADS Bibcode: 2024MNRAS.531..885Z.

[12] Wolfgang Enzi, Riccardo Murgia, Oliver Newton, Simona Vegetti, Carlos Frenk, Matteo Viel, Marius Cautun, Christopher D. Fassnacht, Matt Auger, Giulia Despali, John McKean, Léon V. E. Koopmans, and Mark Lovell. Joint constraints on thermal relic dark matter from strong gravitational lensing, the ly \ensuremath\alpha forest, and milky way satellites. 506(4):5848–5862. _eprint: 2010.13802.

[13] Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, Erik Brynjolfsson, Shyamal Buch, Dallas Card, Rodrigo Castellon, Niladri Chatterji, Annie Chen, Kathleen Creel, Jared Quincy Davis, Dora Demszky, Chris Donahue, Moussa Doumbouya, Esin Durmus, Stefano Ermon, John Etchemendy, Kawin Ethayarajh, Li Fei-Fei, Chelsea Finn, Trevor Gale, Lauren Gillespie, Karan Goel, Noah Goodman, Shelby Grossman, Neel Guha, Tatsunori Hashimoto, Peter Henderson, John Hewitt, Daniel E. Ho, Jenny Hong, Kyle Hsu, Jing Huang, Thomas Icard, Saahil Jain, Dan Jurafsky, Pratyusha Kalluri, Siddharth Karamcheti, Geoff Keeling, Fereshte Khani, Omar Khattab, Pang Wei Koh, Mark Krass, Ranjay Krishna, Rohith Kuditipudi, Ananya Kumar, Faisal Ladhak, Mina Lee, Tony Lee, Jure Leskovec, Isabelle Levent, Xiang Lisa Li, Xuechen Li, Tengyu Ma, Ali Malik, Christopher D. Manning, Suvir Mirchandani, Eric Mitchell, Zanele Munyikwa, Suraj Nair, Avanika Narayan, Deepak Narayanan, Ben Newman, Allen Nie, Juan Carlos Niebles, Hamed Nilforoshan, Julian Nyarko, Giray Ogut, Laurel Orr, Isabel Papadimitriou, Joon Sung Park, Chris Piech, Eva Portelance, Christopher Potts, Aditi Raghunathan, Rob Reich, Hongyu Ren, Frieda Rong, Yusuf Roohani, Camilo Ruiz, Jack Ryan, Christopher Ré, Dorsa Sadigh, Shiori Sagawa, Keshav Santhanam, Andy

Shih, Krishnan Srinivasan, Alex Tamkin, Rohan Taori, Armin W. Thomas, Florian Tramèr, Rose E. Wang, William Wang, Bohan Wu, Jiajun Wu, Yuhuai Wu, Sang Michael Xie, Michihiro Yasunaga, Jiaxuan You, Matei Zaharia, Michael Zhang, Tianyi Zhang, Xikun Zhang, Yuhui Zhang, Lucia Zheng, Kaitlyn Zhou, and Percy Liang. On the opportunities and risks of foundation models. page arXiv:2108.07258. _eprint: 2108.07258.

[14] Michelle Ntampaka, Camille Avestruz, Steven Boada, Joao Caldeira, Jessi Cisewski-Kehe, Rosanne Di Stefano, Cora Dvorkin, August E. Evrard, Arya Farahi, Doug Finkbeiner, Shy Genel, Alyssa Goodman, Andy Goulding, Shirley Ho, Arthur Kosowsky, Paul La Plante, Francois Lanusse, Michelle Lochner, Rachel Mandelbaum, Daisuke Nagai, Jeffrey A. Newman, Brian Nord, J. E. G. Peek, Austin Peel, Barnabas Poczos, Markus Michael Rau, Aneta Siemiginowska, Dougal J. Sutherland, Hy Trac, and Benjamin Wandelt. The role of machine learning in the next decade of cosmology. 51(3):14. _eprint: 1902.10159.

[15] Ana Bonaca, David W. Hogg, Adrian M. Price-Whelan, and Charlie Conroy. The spur and the gap in GD-1: Dynamical evidence for a dark substructure in the milky way halo. 880:38. Publisher: IOP ADS Bibcode: 2019ApJ...880...38B.

[16] Denis Erkal and Vasily Belokurov. Forensics of subhalo-stream encounters: the three phases of gap growth. 450:1136–1149. Publisher: OUP ADS Bibcode: 2015MNRAS.450.1136E.

[17] Euclid Collaboration, Mischa Knabenhans, Joachim Stadel, Stefano Marelli, Doug Potter, Romain Teyssier, Laurent Legrand, Aurel Schneider, Bruno Sudret, Linda Blot, Saeeda Awan, Carlo Burigana, Carla Sofia Carvalho, Hannu Kurki-Suonio, and Gabriele Sirri. Euclid preparation: II. the EUCLIDEMULATOR - a tool to compute the cosmology dependence of the nonlinear matter power spectrum. 484(4):5509–5529. _eprint: 1809.04695.

[18] Raul E. Angulo, Matteo Zennaro, Sergio Contreras, Giovanni Aricò, Marcos Pellejero-Ibañez, and Jens Stücker. The BACCO simulation project: exploiting the full power of large-scale structure for cosmology. 507(4):5869–5881. _eprint: 2004.06245.

[19] Joseph DeRose, Risa H. Wechsler, Jeremy L. Tinker, Matthew R. Becker, Yao-Yuan Mao, Thomas McClintock, Sean McLaughlin, Eduardo Rozo, and Zhongxu Zhai. The AEMULUS project. i. numerical simulations for precision cosmology. 875(1):69. _eprint: 1804.05865.

[20] Katrin Heitmann, Derek Bingham, Earl Lawrence, Steven Bergner, Salman Habib, David Higdon, Adrian Pope, Rahul Biswas, Hal Finkel, Nicholas Frontiere, and Suman Bhattacharya. The mira-titan universe: Precision predictions for dark energy surveys. 820(2):108. _eprint: 1508.02654.

[21] M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. 378:686–707.