Predicting Microbial Ontology and Pathogen Risk from Environmental Metadata with Large Language Models

Hyunwoo Yoo [*1 2] **Gail L. Rosen** [*1]

Abstract

Traditional machine learning models struggle to generalize in microbiome studies where only metadata is available, especially in small-sample settings or across studies with heterogeneous label formats. In this work, we explore the use of large language models (LLMs) to classify microbial samples into ontology categories such as EMPO 3 and related biological labels, as well as to predict pathogen contamination risk, specifically the presence of *E. Coli*, using environmental metadata alone. We evaluate LLMs such as ChatGPT-4o, Claude 3.7 Sonnet, Grok-3, and LLaMA 4 in zero-shot and few-shot settings, comparing their performance against traditional models like Random Forests across multiple real-world datasets. Our results show that LLMs not only outperform baselines in ontology classification, but also demonstrate strong predictive ability for contamination risk, generalizing across sites and metadata distributions. These findings suggest that LLMs can effectively reason over sparse, heterogeneous biological metadata and offer a promising metadata-only approach for environmental microbiology and biosurveillance applications.

1. Introduction

Microbiome classification often relies on sequencing-based taxonomic features. However, such data may be missing or excluded due to cost or design constraints, leaving only environmental metadata for analysis. Environmental metadata, such as material type, biome, sample type, or collection site, is frequently available even in sparse or small-scale studies. Traditional models struggle to generalize on metadata-only settings. This challenge is exacerbated in small-sample settings or when datasets differ in label expressions or structure. These models are limited by their dependence on exact string matches and lack of semantic understanding, which reduces robustness across domains. Even differential abundance methods often yield inconsistent results across datasets, raising reproducibility concerns (Nearing et al., 2022).

Recent advances in large language models (LLMs) present a new opportunity to reason over natural language metadata. Without requiring fine-tuning or manual feature engineering, LLMs can interpret the semantic meaning of metadata fields and align labels across studies. This makes them particularly effective in zero-shot and few-shot classification settings. We adopt these settings to simulate real-world cases where model fine-tuning is not feasible due to sparse or missing sequencing data.

In this work, we investigate whether LLMs can perform two key tasks using environmental metadata alone: classifying microbial samples into ontology categories such as EMPO 3, and predicting pathogen contamination risk, specifically the presence of *E. Coli*. These tasks reflect both ecological characterization and health-related biosurveillance, illustrating the broad utility of LLMs in biological inference.

We evaluate several state-of-the-art LLMs, including ChatGPT-4o (OpenAI et al., 2023), Claude 3.7 Sonnet (Anthropic, 2025), Grok-3 (xAI, 2025), and LLaMA 4 (Meta AI, 2025), under zero-shot and few-shot settings. Their performance is compared to that of traditional models across multiple real-world datasets. Our results show that LLMs not only outperform baselines in ontology classification but also demonstrate strong predictive accuracy for *E. Coli* contamination across varied metadata distributions. These findings highlight the potential of LLMs as metadata-only predictors in microbiome research and environmental health monitoring.

2. Related Work

Traditional microbiome classification relies heavily on taxonomic features derived from high-throughput sequencing data, such as 16S rRNA gene or shotgun metagenomic pro-

[*]Equal contribution [1]Department of Electrical and Computer Engineering, Drexel University, Philadelphia, United States [2]MODULABS, Seoul, Korea. Correspondence to: Hyunwoo Yoo <hty23@drexel.edu>, Gail L. Rosen <glr26@drexel.edu>.

Proceedings of the Workshop on Generative AI for Biology at the 42nd International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025. Copyright 2025 by the author(s).

MODEL	ACCURACY	MACRO PREC.	MACRO REC.	MACRO F1
CHATGPT-4O (ZS)	0.96	0.69	0.75	0.71
CLAUDE 3.7 (ZS)	0.85	0.80	0.83	0.71
GROK-3 (ZS)	0.96	0.69	0.75	0.71
LLAMA 4 (ZS)	0.59	0.44	0.40	0.40
CHATGPT-4O (FS)	0.96	0.69	0.75	0.71
CLAUDE 3.7 (FS)	0.96	0.69	0.75	0.71
GROK-3 (FS)	0.96	0.69	0.75	0.71
LLAMA 4 (FS)	1.00	1.00	1.00	1.00
RANDOM FOREST	0.11	0.03	0.25	0.05

Table 1. Zero-shot and Few-shot inference on Study 15573. In few-shot inference, LLMs receive support examples from Study 1728 (a different domain) as prompt context, enabling cross-study generalization evaluation. LLMs are frozen (no parameter updates), and only Random Forest is trained on Study 1728.

files (Knights et al., 2011; Pasolli et al., 2016). These methods offer accurate microbial characterization but require sequencing resources and may not be feasible in low-cost or metadata-only settings.

In contrast, studies that rely exclusively on environmental metadata for microbiome prediction remain scarce. While there has been growing interest in metadata curation and standardization (Vangay et al., 2021), and recent work highlights reproducibility issues in reusing metadata from 16S studies (Kang et al., 2021), yet few models directly leverage such metadata for ontology or pathogen risk classification.

Recent advancements in large language models (LLMs) have enabled strong zero-shot and few-shot reasoning capabilities across a range of tasks (Brown et al., 2020). In the biomedical domain, LLMs have been used to encode clinical knowledge (Singhal et al., 2023) and integrate multimodal signals such as images and text for medical prediction tasks (Zhang et al., 2024). However, their application to microbiome-related classification using only sparse metadata remains largely unexplored. Moreover, traditional models are highly sensitive to normalization and compositional characteristics of microbiome data, which limits robustness across datasets (Weiss et al., 2017).

Our work bridges this gap by applying foundation models to infer microbial ontology and pathogen contamination risk solely from environmental metadata, demonstrating effective generalization across studies without access to genomic sequences.

3. Method

We use pre-trained large language models (LLMs), including ChatGPT-4o, Claude 3.7 Sonnet, Grok-3, and LLaMA 4, to perform microbiome-related classification and regression tasks using environmental metadata as input. Specifically, we investigate two types of inference problems: (1) classification of microbial samples into standardized ontology categories and related biological labels, and (2) prediction of pathogen contamination risk based on contextual environmental features.

Model	Acc.	Prec.	Rec.	F1
Random Forest	0.47	0.33	0.27	0.30
ChatGPT-4o (ZS)	1.00	1.00	1.00	1.00
Claude 3.7 (ZS)	1.00	1.00	1.00	1.00
Grok-3 (ZS)	1.00	1.00	1.00	1.00

Table 2. Zero-shot inference on Study 1728. Only Random Forest is trained on Study 15573; LLMs are zero-shot prompted.

3.1. Prompting Strategy for LLM-Based Inference

All models are accessed in a frozen setting without fine-tuning. We construct prompts in natural English using structured metadata fields as context. Each prompt queries the model about the most likely label or value associated with a given sample.

Zero-shot classification. In zero-shot settings, a sample x is represented by its metadata fields (e.g., *env material*, *sample type*, *scientific name*, *geo loc name*). We format a multiple-choice prompt that asks whether sample x corresponds to a candidate label $y \in \mathcal{Y}$. The model selects the label with highest likelihood:

$$\hat{y} = \arg\max_{y \in \mathcal{Y}} \text{LLM}_\theta(P(x, y))$$

Few-shot classification. In few-shot settings, the prompt includes a small number of labeled support examples $\mathcal{D}_{\text{support}} = \{(x_i, y_i)\}_{i=1}^{k}$, prepended to the test sample x'. The model selects the most probable label as:

$$\hat{y} = \arg\max_{y \in \mathcal{Y}} \text{LLM}_\theta(P(\mathcal{D}_{\text{support}}, x', y))$$

MODEL	ACCURACY	MACRO PREC.	MACRO REC.	MACRO F1
CHATGPT-4O (ZS)	0.7500	0.7084	0.7210	0.7135
CLAUDE 4SONET (ZS)	0.8036	0.7841	0.7262	0.7441
GROK-3 (ZS)	0.7679	0.7366	0.7670	0.7443
LLAMA 4 (ZS)	0.7143	0.6621	0.6621	0.6621
GEMINI 2.5FLASH (ZS)	0.7857	0.7480	0.7300	0.7375
CHATGPT-4O (FS)	0.8214	0.8202	0.7391	0.7619
CLAUDE 4SONET (FS)	0.7500	0.7044	0.7044	0.7044
GROK-3 (FS)	0.8036	0.8022	0.7097	0.7316
LLAMA 4 (FS)	0.7500	0.7024	0.6878	0.6937
GEMINI 2.5FLASH (FS)	0.7857	0.7825	0.6802	0.6995

Table 3. Zero-shot and Few-shot *E. Coli* level binary prediction on the 2005 Huntington Beach dataset. In the few-shot setting, LLMs receive support examples from 2006 samples, enabling evaluation of cross-year generalization across temporally shifted metadata distributions.

To evaluate cross-study generalization, the support examples are drawn from a different dataset than the test set (e.g., Study 1728 as support for Study 15573 and vice versa), and no model parameters are updated. Similarly, in the *E. Coli* contamination prediction task, we evaluate cross-year generalization by using 2005 Huntington Beach data for testing and 2006 samples as support examples in few-shot prompting.

Binary contamination prediction. For *E. Coli* contamination risk, we cast the task as binary classification using a prompt that asks whether the microbial contamination level exceeds a threshold such as 126 CFU/100mL as per EPA guidelines(United States Environmental Protection Agency, 2012; 2021). We evaluate LLM responses against ground-truth labels derived from measured *E. Coli* levels.

Regression prediction. We additionally evaluate whether LLMs can perform numeric regression by prompting for the predicted concentration of *E. Coli* based on a sample's environmental metadata. The model's free-text numeric output is extracted and compared to ground-truth concentrations using standard regression metrics such as MAE and R^2.

3.2. Prompt Format and Standardization

All prompts follow a fixed structure, include consistent field ordering, and avoid ambiguous phrasing. We use prompt templates that include direct natural language questions (e.g., "Given the following metadata, what is the most likely EMPO 3 category for this sample?"). For regression tasks, prompts explicitly ask for a numerical estimate (e.g., "Estimate the *E. Coli* concentration (in CFU/100mL) based on the following metadata."). Examples of all prompt templates are included in Appendix C.

4. Experiments

We evaluate large language models (LLMs) on two tasks using only environmental metadata: (1) microbial context classification and (2) *E. Coli* contamination risk prediction. We consider zero-shot and few-shot prompting with no fine-tuning, and compare LLMs to traditional machine learning models such as Random Forest and XGBoost. Details of environmental metadata are in Appendix A

Model	Acc.	Prec.	Rec.	F1
ChatGPT-4o (ZS)	0.6721	0.6482	0.6424	0.6445
Claude 4sonet (ZS)	0.7049	0.6872	0.6602	0.6652
Grok-3 (ZS)	0.7049	0.6889	0.6945	0.6909
LLaMA 4 (ZS)	0.7377	0.7398	0.7551	0.7342
Gemini 2.5flash (ZS)	NA	NA	NA	NA

Table 4. Zero-shot *E. Coli* binary prediction on 2006 Huntington Beach data.

4.1. Microbial Context Classification

We begin with EMPO 3 classification using metadata fields such as 'env material' and 'sample type' in study 1728 and 15573(Hewson et al., 2022; Baum & Ackerman, 2022). Table 2 shows that all LLMs achieve perfect accuracy on Study 1728 in zero-shot prompting, while Random Forest trained on Study 15573 fails (47%). This highlights LLMs' ability to generalize across studies using metadata alone. On Study 15573, which contains a more diverse set of environments and labels, zero-shot performance remains high for ChatGPT-4o and Grok-3 (96% accuracy), while Claude 3.7 trails slightly (85%) and LLaMA 4 struggles at 59% (Table 1). This indicates LLMs' ability to perform semantic alignment between metadata and ontology labels across domains. When given few-shot prompts using Study 1728 samples (Table 1), all models improve or maintain strong performance, with LLaMA 4 jumping to 100% accuracy. Additional results on 'sample_type' and 'scientific_name' classification appear in Tables 8–10 in Appendix B.

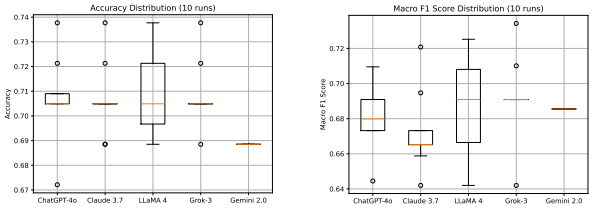
4.2. *E. Coli* Contamination Prediction

Model	Acc.	Prec.	Rec.	F1
ChatGPT-4o (FS)	NA	NA	NA	NA
Claude 4sonet (FS)	0.6885	0.6866	0.6985	0.6831
Grok-3 (FS)	0.7705	0.7567	0.7643	0.7596
LLaMA 4 (FS)	NA	NA	NA	NA
Gemini 2.5flash (FS)	0.8033	0.7906	0.7906	0.7906

Table 5. Few-shot *E. Coli* binary prediction on 2006 Huntington Beach data.

We next evaluate binary classification of *E. Coli* risk (above/below regulatory threshold). As shown in the table 1 on 2005 Huntington Beach data, zero-shot prompting shows strong performance. For instance, Claude 4 Sonnet reaches 80.4% accuracy and 0.7441 macro F1, followed closely by Grok-3 and Gemini 2.5Flash. Few-shot prompting further boosts ChatGPT-4o to 82.1% accuracy and 0.7619 F1, outperforming all others. This setup mirrors realistic scenarios where prior year data is used to inform current year risk, enabling cross-year generalization without sequence data.

On 2006 data(Francy et al., 2021), LLaMA 4 surprisingly leads in zero-shot performance (73.8% accuracy), while Claude and Grok-3 also perform well (Table 4). Gemini fails to return predictions. In the few-shot setting (Table 5), Gemini achieves the highest overall score (80.3% accuracy), followed by Grok-3 and Claude. This suggests LLMs generalize reasonably well across yearly shifts in metadata and environment.



(a) Accuracy distribution **(b)** F1 Score distribution

Figure 1. Accuracy and F1 Score distributions across 10 repeated trials for each LLM in *E. Coli* level binary prediction using the 2005 Huntington Beach dataset.

4.3. Robustness and Regression Results

We repeat binary classification over 10 random trials on the 2005 dataset. Figure 1 and Table 11 in Appendix B show that Grok-3 and Claude 4 Sonnet exhibit the most stable performance. LLaMA 4 displays high variance due to limited valid outputs (N=3), and Gemini's single-run result suggests limited robustness.

For numeric prediction of *E. Coli* concentration, zero-shot regression fails for most LLMs (Table 6). Claude 4 Sonnet returns highly unstable estimates with $R^2 = -7.71$. In contrast, in the few-shot setting (Table 7), Claude achieves $R^2 = 0.3946$, outperforming all traditional models, including Random Forest (0.3261). However, other LLMs such as ChatGPT-4o and Grok-3 still underperform, and variance remains high. This suggests that while LLMs are effective for classification, they are still unreliable for quantitative microbial estimation.

Model	MAE	RMSE	MSE	R^2
ChatGPT-4o	287.56	577.70	333733.36	-0.0673
Claude 4sonet	347.20	1650.71	2724857.43	-7.7145
LLaMA 4	N/A	N/A	N/A	N/A
Grok-3	261.32	564.57	318737.63	-0.0194
Gemini 2.5flash	N/A	N/A	N/A	N/A

Table 6. Zero-shot regression performance on 2006 Huntington beach *E. Coli* data (MAE, RMSE, MSE, R^2).

Model	MAE	RMSE	MSE	R^2
ChatGPT-4o	259.14	612.07	374632.53	-0.1981
Claude 4sonet	190.96	435.10	189310.24	0.3946
LLaMA 4	N/A	N/A	N/A	N/A
Grok-3	190.67	493.97	244002.06	0.2196
Gemini 2.5flash	232.34	478.71	229159.20	0.2671
Random Forest	224.69	459.03	210709.20	0.3261
XGBoost	251.89	542.17	293954.79	0.0599
Logistic Reg.	220.76	524.28	274872.34	0.1209

Table 7. Few-shot regression performance on 2006 Huntington beach *E. Coli* data (MAE, RMSE, MSE, R^2).

5. Results and Discussion

LLMs consistently outperform traditional models in ontology classification, even when tested on out-of-distribution samples. Cross-study generalization is especially strong for ChatGPT-4o and Grok-3.

While zero-shot prompting already achieves strong binary classification of *E. Coli* presence, few-shot examples further enhance accuracy and F1. Claude and Grok-3 are the most robust models across years.

Despite some success from Claude 4 Sonnet in few-shot settings, LLMs are not yet competitive with traditional regressors for numeric estimation. Prediction variance and output formatting remain issues.

6. Conclusion

We show that large language models (LLMs) can accurately classify microbial ontology labels and predict pathogen contamination risk using only environmental metadata. Across diverse datasets and tasks, LLMs demonstrate strong zero-shot and few-shot performance, often outperforming traditional models without requiring fine-tuning. These results highlight the capacity of LLMs to semantically reason over heterogeneous metadata fields, enabling generaliza-

tion across studies and label variations. While LLM-based regression remains less reliable, classification results suggest that foundation models offer a promising sequence-free approach for microbiome analysis and environmental biosurveillance.

Acknowledgements

This work was supported in part by the National Science Foundation (NSF) under Grant Number 2107108 and Brian Impact Foundation, a non-profit organization dedicated to the advancement of science and technology for all.

References

Anthropic. Claude 3.7 sonnet and claude code. https://www.anthropic.com/news/claude-3-7-sonnet, February 2025. Accessed: 2025-05-25.

Baum, M. and Ackerman, G. Baum asphalt 1st submission. https://qiita.ucsd.edu/study/description/1728, 2022. Qiita Study ID: 1728, Samples: 17, PI: Marc Baum (Oak Crest Laboratory), EBI Accession: not submitted.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., and Amodei, D. Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33:1877–1901, 2020.

Francy, D. S., Brady, A. M., and Zimmerman, T. M. Data for multiple linear regression models for estimating *Escherichia coli* (e. coli) concentrations or the probability of exceeding the bathing-water standard at recreational sites in ohio and pennsylvania as part of the great lakes nowcast, 2019, 2021. URL https://doi.org/10.5066/P9Y9O1YJ. USGS Data Release.

Hewson, I., Cedeno, B., and collaborators. Detection of the *diadema antillarum* scuticociliatosis *philaster* clade on sympatric metazoa, plankton, and abiotic surfaces and assessment for its potential reemergence. https://qiita.ucsd.edu/study/description/15573, 2022. Qiita Study ID: 15573, EBI Accession: ERP161028, PI: Ian Hewson (Cornell University), Lab Contact: Brayan (UPR School of Medicine).

Kang, X., Deng, D. M., Crielaard, W., and Brandt, B. W. Reprocessing 16s rrna gene amplicon sequencing studies: (meta)data issues, robustness, and reproducibility. *Frontiers in Cellular and Infection Microbiology*, 11:720637, 2021. doi: 10.3389/fcimb.2021.720637.

Knights, D., Costello, E. K., and Knight, R. Supervised classification of human microbiota. *FEMS microbiology reviews*, 35(2):343–359, 2011.

Meta AI. The llama 4 herd: The beginning of a new era of natively multimodal ai innovation. https://ai.meta.com/blog/llama-4-multimodal-intelligence/, April 2025. Accessed: 2025-05-25.

Nearing, J. T., Douglas, G. M., Hayes, M. G., MacDonald, J., Desai, D. K., Allward, N., Jones, C. M. A., Wright, R. J., Dhanani, A. S., Comeau, A. M., and Langille, M. G. I. Microbiome differential abundance methods produce different results across 38 datasets. *Nature Communications*, 13:342, 2022. doi: 10.1038/s41467-022-28034-z.

OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao, H., Bavarian, M., Belgum, J., Bello, I., Berdine, J., Bernadett-Shapiro, G., Berner, C., Bogdonoff, L., Boiko, O., Boyd, M., Brakman, A.-L., Brockman, G., Brooks, T., Brundage, M., Button, K., Cai, T., Campbell, R., Cann, A., Carey, B., Carlson, C., Carmichael, R., Chan, B., Chang, C., Chantzis, F., Chen, D., Chen, S., Chen, R., Chen, J., Chen, M., Chess, B., Cho, C., Chu, C., Chung, H. W., Cummings, D., Currier, J., Dai, Y., Decareaux, C., Degry, T., Deutsch, N., Deville, D., Dhar, A., Dohan, D., Dowling, S., Dunning, S., Ecoffet, A., Eleti, A., Eloundou, T., Farhi, D., Fedus, L., Felix, N., Fishman, S. P., Forte, J., Fulford, I., Gao, L., Georges, E., Gibson, C., Goel, V., Gogineni, T., Goh, G., Gontijo-Lopes, R., Gordon, J., Grafstein, M., Gray, S., Greene, R., Gross, J., Gu, S. S., Guo, Y., Hallacy, C., Han, J., Harris, J., He, Y., Heaton, M., Heidecke, J., Hesse, C., Hickey, A., Hickey, W., Hoeschele, P., Houghton, B., Hsu, K., Hu, S., Hu, X., Huizinga, J., Jain, S., Jain, S., et al. Gpt-4 technical report. https://arxiv.org/abs/2303.08774, March 2023. arXiv:2303.08774 [cs.CL].

Pasolli, E., Truong, D. T., Malik, F., Waldron, L., and Segata, N. Machine learning meta-analysis of large metagenomic datasets: tools and biological insights. *PLOS Computational Biology*, 12(7):e1004977, 2016.

Singhal, K., Azizi, S., Tu, T., Mahdavi, S. S., Wei, J., Chung, H. W., Scales, N., Tanwani, A., Cole-Lewis, H., Pfohl, S., Payne, P., Seneviratne, M., Gamble, P., Kelly, C., Babiker, A., Schärli, N., Chowdhery, A., Mansfield, P., Demner-Fushman, D., Agüera y Arcas, B., Webster, D., Corrado, G. S., Matias, Y., Chou, K., Gottweis, J., Tomasev, N., Liu, Y., Rajkomar, A., Barral, J., Semturs, C., Karthikesalingam, A., and Natarajan, V. Large language models encode clinical knowledge. *Nature*, 620:172–180, 2023.

United States Environmental Protection Agency. Recreational water quality criteria, 2012. URL https://www.epa.gov/sites/default/files/2015-10/documents/rwqc2012.pdf.

United States Environmental Protection Agency. E. coli parameter factsheet, 2021. URL https://www.epa.gov/system/files/documents/2021-07/parameter-factsheet_e.-coli.pdf.

Vangay, P., Burgin, J., Johnston, A., Beck, K. L., Berrios, D. C., Blumberg, K., Canon, S., Chain, P., Chandonia, J.-M., Christianson, D., Costes, S. V., Damerow, J., Duncan, W. D., Dundore-Arias, J. P., Fagnan, K., Galazka, J. M., Gibbons, S. M., Hays, D., Hervey, J., Hu, B., Hurwitz, B. L., Jaiswal, P., Joachimiak, M. P., Kinkel, L., Ladau, J., Martin, S. L., McCue, L. A., Miller, K., Mouncey, N., Mungall, C., Pafilis, E., Reddy, T. B. K., Richardson, L., Roux, S., Schriml, L. M., Shaffer, J. P., Sundaramurthi, J. C., Thompson, L. R., Timme, R. E., Zheng, J., Wood-Charlson, E. M., and Eloe-Fadrosh, E. A. Microbiome metadata standards: Report of the national microbiome data collaborative's workshop and follow-on activities. *mSystems*, 6(1), 2021. doi: 10.1128/mSystems.01194-20.

Weiss, S., Xu, Z. Z., Peddada, S. D., Amir, A., Bittinger, K., Gonzalez, A., Lozupone, C., Zaneveld, J. R., Vázquez-Baeza, Y., Birmingham, A., Hyde, E. R., and Knight, R. Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome*, 5:27, 2017. doi: 10.1186/s40168-017-0237-y.

xAI. Grok 3 Beta — The Age of Reasoning Agents. https://x.ai/news/grok-3, February 2025. Accessed: 2025-05-25.

Zhang, K., Zhou, R., Adhikarla, E., Yan, Z., Liu, Y., Yu, J., Liu, Z., Chen, X., Davison, B. D., Ren, H., Huang, J., Chen, C., Zhou, Y., Fu, S., Liu, W., Liu, T., Li, X., Chen, Y., He, L., Zou, J., Li, Q., Liu, H., and Sun, L. A generalist vision-language foundation model for diverse biomedical tasks. *Nature Medicine*, 30(11):3129–3141, Nov 2024. doi: 10.1038/s41591-024-03185-2. Epub 2024 Aug 7.

A. Dataset Details

Study 1728 (Baum asphalt 1st submission)(Baum & Ackerman, 2022) involved bacterial metagenomic sequencing of samples collected from various asphalt sites as well as nearby water and soil locations. Additionally, three samples (2H, 2L, and 2O) were cultured in LB broth before being included in the overall analysis. This study comprises a total of 17 samples, categorized into two EMPO 3 labels: 10 samples as Solid (non-saline) and 7 samples as Aqueous (non-saline).

The metadata for this study includes five primary features: env_material, env_biome, env_feature, sample_type, and scientific_name. The feature values are relatively simple and consistent. For env_material, the categories include water, soil, and anthropogenic environmental material. The env_biome is exclusively desert biome. All samples are associated with the env_feature road. The sample_type includes water filter, water and LB broth, soil, and asphalt. The scientific_name consists of freshwater metagenome, soil metagenome, and outdoor metagenome. This study does not include geographic information such as geo_loc_name, and the overall structure of metadata is constrained and minimal.

Study 15573 (Detection of the Diadema antillarum scuticociliatosis Philaster clade on sympatric metazoa, plankton, and abiotic surfaces and assessment for its potential reemergence) (Hewson et al., 2022) is an investigation into the distribution and persistence of a scuticociliate pathogen (DaScPc) that caused mass mortality of long-spined sea urchins throughout the eastern Caribbean in 2022. A total of 27 samples were analyzed, collected from coral species, turf algae, and sponges, particularly near the original outbreak site. The EMPO 3 label distribution includes 17 samples as Animal (saline), 6 as Plant (saline), 3 as Solid (non-saline), and 1 as Aqueous (saline), indicating greater label diversity compared to Study 1728.

The metadata for Study 15573 contains six features: env_material, env_biome, env_feature, sample_type, scientific_name, and geo_loc_name. The env_material includes organic material and anthropogenic environmental material. The env_biome consists of marine biome and urban biome. The env_feature spans coral reef, animal-associated habitat, plant-associated habitat, anthropogenic environmental feature, and research facility. The sample_type includes categories such as Turf Algae, coral, hydrozoans, sponge, Boat Hull, Mangrove Leaf, and control swab. The scientific_name categories include algae metagenome, coral metagenome, sponge metagenome, and plant metagenome. The geo_loc_name feature provides geographic labels such as US Virgin Islands and Aruba, offering richer contextual information.

Overall, Study 15573 features more complex metadata with greater categorical diversity across multiple fields and includes detailed geographic context. In comparison, Study 1728 is more constrained in terms of both label space and feature variety. These differences contribute to substantial domain shift, which presents a challenge for conventional machine learning models trained on one study and evaluated on the other.

We also utilize a subset of the dataset released by the U.S. Geological Survey as part of the Great Lakes NowCast project (Francy et al., 2021), specifically the calibration data collected at Huntington Beach, Ohio, during the 2005 and 2006 recreational seasons. The dataset contains seven variables. Date indicates the sampling date. EcoliAve_CFU represents the average concentration of *Escherichia coli* (*E. Coli*), measured in colony-forming units (CFU). Lake_Temp_C denotes the lake water temperature in degrees Celsius, while Lake_Turb_NTRU captures the turbidity of the lake, measured in Nephelometric Turbidity Units (NTU). WaveHt_Ft indicates the height of waves in feet at the time of sampling. LL_PreDay refers to the change in lake level compared to the previous day. Lastly, AirportRain48W_in measures the cumulative rainfall over the past 48 hours as recorded at a nearby airport, expressed in inches.

B. Addtional Results

Model	Accuracy	Macro Prec.	Macro Rec.	Macro F1
ChatGPT-4o (ZS)	1.0	1.0	1.0	1.0
Claude 3.7 (ZS)	1.0	1.0	1.0	1.0
Grok-3 (ZS)	1.0	1.0	1.0	1.0
LLaMA 4 (ZS)	1.0	1.0	1.0	1.0

Table 8. Zero-shot sample_type prediction on Study 15573.

We present additional zero-shot classification results on Study 15573 to further evaluate the robustness of LLMs under different label and feature configurations.

MODEL	ACCURACY	MACRO PREC.	MACRO REC.	MACRO F1
CHATGPT-4O (ZS)	1.0	1.0	1.0	1.0
CLAUDE 3.7 (ZS)	1.0	1.0	1.0	1.0
GROK-3 (ZS)	1.0	1.0	1.0	1.0
LLAMA 4 (ZS)	1.0	1.0	1.0	1.0

Table 9. Zero-shot scientific_name prediction on Study 15573.

MODEL	ACCURACY	MACRO PREC.	MACRO REC.	MACRO F1
CHATGPT-4O (ZS)	0.4074	0.3824	0.5000	0.4091
CLAUDE 3.7 (ZS)	0.4074	0.3824	0.5000	0.4091
GROK-3 (ZS)	0.4074	0.3824	0.5000	0.4091
LLAMA 4 (ZS)	0.4074	0.3824	0.5000	0.4091

Table 10. Zero-shot scientific_name prediction on Study 15573 without sample_type feature.

Sample type prediction. As shown in Table 8, all LLMs achieved perfect accuracy when predicting the sample_type field using available metadata. This confirms that even simple categorical labels can be reliably inferred by LLMs in zero-shot settings without fine-tuning.

Scientific name prediction. Table 9 reports results for predicting scientific_name labels. Again, all LLMs achieved perfect accuracy, indicating strong semantic reasoning over metadata features that are biologically descriptive.

Impact of removing sample_type. To assess the importance of the features, we removed the sample_type field and repeated the scientific_name classification. As seen in Table 10, accuracy drops significantly to 40.7% for all models, with macro F1 around 0.41. This suggests that sample_type is a critical contextual cue for correctly identifying scientific labels and that LLMs are sensitive to the availability of relevant metadata features.

MODEL	METRIC	MEAN	STD	MIN	MAX	VALID N
CHATGPT-4O	ACCURACY	0.7070	0.0173	0.6721	0.7377	8
	PRECISION	0.6892	0.0204	0.6482	0.7231	8
	RECALL	0.6779	0.0182	0.6424	0.7037	8
	F1	0.6802	0.0175	0.6445	0.7095	8
CLAUDE 4 SONET	ACCURACY	0.7067	0.0143	0.6885	0.7377	9
	PRECISION	0.6877	0.0157	0.6659	0.7208	9
	RECALL	0.6693	0.0223	0.6384	0.7208	9
	F1	0.6731	0.0213	0.6420	0.7208	9
LLAMA 4	ACCURACY	0.7104	0.0205	0.6885	0.7377	3
	PRECISION	0.6933	0.0225	0.6681	0.7228	3
	RECALL	0.6874	0.0375	0.6384	0.7294	3
	F1	0.6860	0.0341	0.6420	0.7252	3
GROK-3	ACCURACY	0.7085	0.0129	0.6885	0.7377	9
	PRECISION	0.6943	0.0186	0.6681	0.7398	9
	RECALL	0.6974	0.0283	0.6384	0.7551	9
	F1	0.6924	0.0226	0.6420	0.7342	9
GEMINI 2.5 FLASH	ACCURACY	0.6885	0.0000	0.6885	0.6885	1
	PRECISION	0.6950	0.0000	0.6950	0.6950	1
	RECALL	0.7071	0.0000	0.7071	0.7071	1
	F1	0.6855	0.0000	0.6855	0.6855	1

Table 11. Detailed statistics of LLM performance across 10 repeated runs (or fewer if response failures occurred) in Zero-shot *E. Coli* level binary prediction based on the Data collected in 2006 from Huntington Beach, Ohio. We report the mean, standard deviation, minimum, and maximum values, as well as the number of valid runs (Valid N) for each metric.

To assess the robustness of each LLM, we repeated the zero-shot *E. Coli* classification task 10 times using the 2006 Huntington Beach dataset. Table 11 summarizes the mean, standard deviation, and range of evaluation metrics across repeated runs. Notably, Claude 4 Sonnet and Grok-3 exhibited the most consistent performance, while models like LLaMA 4 and Gemini 2.5 Flash showed higher variance or failed to return valid predictions in multiple runs. These findings highlight the varying stability and reliability of LLMs in structured binary classification tasks.

C. Prompt

C.1. Zero-shot EMPO_3 classification prompt example

```
Could you predict the "empo_3" values below(which is now in '?') with Animal (saline) or
Plant (saline) or Solid (non-saline) or Aqueous (saline) ?


env_material env_biome env_feature sample_type scientific_name geo_loc_name empo_3
0 organic material marine biome coral reef Turf Algae algae metagenome US Virgin Islands ?
1 organic material marine biome animal-associated habitat coral coral metagenome Aruba ?
2 organic material marine biome animal-associated habitat coral coral metagenome US Virgin Islands ?
3 organic material marine biome plant-associated habitat Mangrove Leaf plant metagenome US Virgin Islands ?
4 organic material marine biome coral reef Turf Algae algae metagenome US Virgin Islands ?
5 organic material marine biome animal-associated habitat coral coral metagenome Aruba ?
6 organic material marine biome animal-associated habitat hydrozoans hydrozoan metagenome Aruba ?
7 organic material marine biome animal-associated habitat coral coral metagenome US Virgin Islands ?
8 organic material marine biome animal-associated habitat sponge sponge metagenome US Virgin Islands ?
9 organic material marine biome animal-associated habitat sponge sponge metagenome US Virgin Islands ?
10 organic material marine biome animal-associated habitat coral coral metagenome Aruba ?
11 organic material marine biome coral reef Turf Algae algae metagenome US Virgin Islands ?
12 organic material marine biome animal-associated habitat sponge sponge metagenome US Virgin Islands ?
13 organic material marine biome animal-associated habitat coral coral metagenome US Virgin Islands ?
14 organic material marine biome anthrogenic environmental feature Boat Hull metagenome US Virgin Islands ?
15 organic material marine biome animal-associated habitat sponge sponge metagenome US Virgin Islands ?
16 organic material marine biome animal-associated habitat coral coral metagenome US Virgin Islands ?
17 organic material marine biome animal-associated habitat coral coral metagenome US Virgin Islands ?
18 anthropogenic environmental material urban biome research facility control swab metagenome US Virgin Islands ?
19 anthropogenic environmental material urban biome research facility control swab metagenome US Virgin Islands ?
20 anthropogenic environmental material urban biome research facility control swab metagenome US Virgin Islands ?
21 organic material marine biome plant-associated habitat Mangrove Leaf plant metagenome US Virgin Islands ?
22 organic material marine biome animal-associated habitat coral coral metagenome US Virgin Islands ?
23 organic material marine biome animal-associated habitat coral coral metagenome US Virgin Islands ?
24 organic material marine biome coral reef Turf Algae algae metagenome Aruba ?
25 organic material marine biome animal-associated habitat coral coral metagenome Aruba ?
26 organic material marine biome animal-associated habitat sponge sponge metagenome US Virgin Islands ?
```

C.2. Few-shot EMPO_3 classification prompt example

```
Based on this study:

env_material env_biome env_feature sample_type scientific_name empo_3
0 water desert biome road water filter freshwater metagenome Aqueous (non-saline)
1 water desert biome road water filter freshwater metagenome Aqueous (non-saline)
2 soil desert biome road soil soil metagenome Solid (non-saline)
3 anthropogenic environmental material desert biome road asphalt outdoor metagenome Solid (non-saline)
4 anthropogenic environmental material desert biome road asphalt outdoor metagenome Solid (non-saline)
5 anthropogenic environmental material desert biome road asphalt outdoor metagenome Solid (non-saline)
6 anthropogenic environmental material desert biome road asphalt outdoor metagenome Solid (non-saline)
7 water desert biome road water and LB broth freshwater metagenome Aqueous (non-saline)
8 soil desert biome road soil soil metagenome Solid (non-saline)
9 water desert biome road water and LB broth freshwater metagenome Aqueous (non-saline)
10 anthropogenic environmental material desert biome road asphalt outdoor metagenome Solid (non-saline)
11 anthropogenic environmental material desert biome road asphalt outdoor metagenome Solid (non-saline)
12 water desert biome road water and LB broth freshwater metagenome Aqueous (non-saline)
13 anthropogenic environmental material desert biome road asphalt outdoor metagenome Solid (non-saline)
14 anthropogenic environmental material desert biome road asphalt outdoor metagenome Solid (non-saline)
15 water desert biome road water filter freshwater metagenome Aqueous (non-saline)
16 water desert biome road water filter freshwater metagenome Aqueous (non-saline)


env_material env_biome env_feature sample_type scientific_name geo_loc_name empo_3
0 organic material marine biome coral reef Turf Algae algae metagenome US Virgin Islands ?
1 organic material marine biome animal-associated habitat coral coral metagenome Aruba ?
2 organic material marine biome animal-associated habitat coral coral metagenome US Virgin Islands ?
3 organic material marine biome plant-associated habitat Mangrove Leaf plant metagenome US Virgin Islands ?
4 organic material marine biome coral reef Turf Algae algae metagenome US Virgin Islands ?
5 organic material marine biome animal-associated habitat coral coral metagenome Aruba ?
6 organic material marine biome animal-associated habitat hydrozoans hydrozoan metagenome Aruba ?
7 organic material marine biome animal-associated habitat coral coral metagenome US Virgin Islands ?
8 organic material marine biome animal-associated habitat sponge sponge metagenome US Virgin Islands ?
9 organic material marine biome animal-associated habitat sponge sponge metagenome US Virgin Islands ?
10 organic material marine biome animal-associated habitat coral coral metagenome Aruba ?
11 organic material marine biome coral reef Turf Algae algae metagenome US Virgin Islands ?
12 organic material marine biome animal-associated habitat sponge sponge metagenome US Virgin Islands ?
13 organic material marine biome animal-associated habitat coral coral metagenome US Virgin Islands ?
```

```
14 organic material marine biome anthrogenic environmental feature Boat Hull metagenome US Virgin Islands ?
15 organic material marine biome animal-associated habitat sponge sponge metagenome US Virgin Islands ?
16 organic material marine biome animal-associated habitat coral coral metagenome US Virgin Islands ?
17 organic material marine biome animal-associated habitat coral coral metagenome US Virgin Islands ?
18 anthropogenic environmental material urban biome research facility control swab metagenome US Virgin Islands ?
19 anthropogenic environmental material urban biome research facility control swab metagenome US Virgin Islands ?
20 anthropogenic environmental material urban biome research facility control swab metagenome US Virgin Islands ?
21 organic material marine biome plant-associated habitat Mangrove Leaf plant metagenome US Virgin Islands ?
22 organic material marine biome animal-associated habitat coral coral metagenome US Virgin Islands ?
23 organic material marine biome animal-associated habitat coral coral metagenome US Virgin Islands ?
24 organic material marine biome coral reef Turf Algae algae metagenome Aruba ?
25 organic material marine biome animal-associated habitat coral coral metagenome Aruba ?
26 organic material marine biome animal-associated habitat sponge sponge metagenome US Virgin Islands ?

Could you predict the "empo_3" values(which is now in '?') with Animal (saline) or Plant (saline) or Solid (non-saline) or Aqueous (saline) ?
```

C.3. Zero-shot *E. Coli* risk binary classification prompt example

Could you classify the following 56 rows using the Ecoli_binary label, where 1 indicates non-safe freshwater and 0 indicates safe freshwater?

Date	Lake_Temp_C	Lake_Turb_NTRU	WaveHt_Ft	LL_PreDay	AirportRain48W_in
2005-05-25	13.3	58.0	1.00	-0.099	0.0
2005-05-26	14.4	11.5	1.00	-0.256	0.0
2005-05-31	17.8	3.2	0.00	0.027	0.1
2005-06-01	17.8	3.5	0.50	0.045	0.0
2005-06-02	17.2	8.4	1.00	-0.062	0.0
2005-06-06	18.3	7.1	0.33	0.184	0.1
2005-06-07	20.0	9.3	0.50	-0.328	0.1
2005-06-08	20.0	3.0	0.00	0.108	0.0
2005-06-09	23.3	1.9	0.00	0.026	0.0
2005-06-13	24.4	2.9	0.00	0.079	0.1
2005-06-14	21.7	5.2	1.00	-0.069	0.3
2005-06-15	21.1	18.5	1.00	-0.148	1.1
2005-06-16	20.0	54.5	1.50	0.184	1.3
2005-06-20	22.2	32.5	0.50	-0.217	0.0
2005-06-21	21.7	26.0	0.50	-0.029	0.0
2005-06-22	22.8	23.0	0.50	0.029	0.2
2005-06-23	22.8	25.5	0.50	0.027	0.1
2005-06-27	24.4	9.8	0.00	-0.075	0.0
2005-06-28	23.9	4.2	0.00	0.049	0.0
2005-06-29	24.4	7.0	0.50	0.072	0.3
2005-06-30	24.4	3.6	0.00	-0.144	0.3
2005-07-05	24.4	6.2	1.00	0.069	0.0
2005-07-06	23.3	16.0	1.50	0.036	0.0
2005-07-07	24.4	12.0	1.00	-0.099	0.0
2005-07-11	24.4	3.8	0.50	0.004	0.0
2005-07-12	24.4	3.0	0.50	-0.073	0.0
2005-07-13	25.6	4.7	0.50	-0.009	0.0
2005-07-14	24.4	4.4	0.50	0.006	0.1
2005-07-18	25.0	4.4	0.50	-0.014	0.2
2005-07-19	24.4	4.3	0.50	0.050	0.4
2005-07-20	26.1	1.7	0.50	0.013	0.2
2005-07-21	23.9	6.1	0.50	-0.138	1.5
2005-07-25	25.6	8.3	0.50	-0.053	0.5
2005-07-26	25.6	2.3	0.50	-0.236	0.3
2005-07-27	24.4	62.5	5.00	0.857	2.9
2005-07-28	22.2	14.7	2.00	-0.584	1.9
2005-08-01	26.1	4.7	0.00	-0.003	0.0

2005-08-02	26.7	3.0	0.50	-0.053	0.0
2005-08-03	27.2	2.0	0.00	0.007	0.0
2005-08-04	26.1	1.6	0.50	-0.089	0.0
2005-08-08	26.7	1.7	1.00	-0.052	0.0
2005-08-09	26.7	1.5	0.00	-0.138	0.0
2005-08-10	26.1	1.9	0.00	0.023	0.0
2005-08-11	26.1	8.0	2.00	0.253	0.3
2005-08-15	25.6	13.5	2.00	0.062	0.3
2005-08-16	24.4	4.8	0.33	-0.095	0.0
2005-08-17	25.6	2.3	0.17	-0.118	0.0
2005-08-18	25.0	2.3	0.66	-0.040	0.0
2005-08-22	24.4	86.5	3.00	0.076	3.6
2005-08-23	24.4	28.5	4.00	0.023	0.1
2005-08-24	24.4	10.2	2.50	-0.050	0.1
2005-08-25	24.4	10.0	2.00	-0.032	0.0
2005-08-29	23.3	2.7	0.50	-0.282	1.2
2005-08-30	23.3	2.4	0.50	0.207	0.0
2005-09-01	22.8	52.0	2.33	-0.167	2.2
2005-09-06	21.7	21.5	1.00	-0.229	0.0

Please show the result in Python list format.

C.4. Few-shot *E. Coli* risk binary classification prompt example

Based on below study:

Date	Lake_Temp_C	Lake_Turb_NTRU	WaveHt_Ft	LL_PreDay	AirportRain48W_in	Ecoli_binary
2006-06-01	20.0	3.9	0.50	0.040		0.7
2006-06-05	18.3	21.8	0.50	-0.010		0.2
2006-06-06	18.9	10.0	0.00	0.029		0.0
2006-06-07	21.1	3.7	0.00	0.030		0.0
2006-06-08	20.0	13.5	0.50	0.010		0.0
2006-06-12	18.3	41.1	2.50	0.017		0.0
2006-06-13	18.3	27.5	1.50	-0.158		0.0
2006-06-14	18.9	6.7	0.50	-0.046		0.0
2006-06-15	18.9	22.1	1.00	-0.003		0.0
2006-06-19	20.0	11.9	0.50	0.217		1.6
2006-06-20	20.6	7.3	1.00	0.092		1.9
2006-06-21	21.7	3.8	0.00	-0.325		0.6
2006-06-22	21.1	17.2	0.00	0.203		4.6
2006-06-23	20.0	5.6	0.00	0.105		3.7
2006-06-24	21.1	35.0	2.50	-0.062		0.8
2006-06-25	22.8	9.8	0.00	0.148		0.1
2006-06-26	21.1	11.2	0.50	-0.017		0.0
2006-06-27	21.1	14.5	0.00	0.105		0.7
2006-07-01	22.8	5.0	0.00	-0.053		0.0
2006-07-02	22.8	5.0	0.00	-0.269		0.0
2006-07-03	23.3	5.9	0.50	0.158		0.2
2006-07-04	21.7	9.1	0.00	0.095		1.1
2006-07-05	21.1	57.5	2.50	0.128		0.6
2006-07-11	23.9	7.6	0.50	0.214		0.2

2006-07-12	22.8	4.7	0.00	-0.161	1.3
2006-07-13	22.2	16.7	1.00	0.272	0.8
2006-07-14	22.8	4.2	0.50	-0.085	0.1
2006-07-15	23.3	19.4	0.00	0.065	1.0
2006-07-16	23.9	4.0	0.00	-0.121	0.5
2006-07-18	26.1	5.8	0.50	-0.174	0.0
2006-07-22	25.6	23.1	3.00	0.180	0.6
2006-07-24	25.6	5.1	0.00	-0.124	0.1
2006-07-25	25.0	3.8	0.50	-0.142	0.0
2006-07-26	25.6	2.9	0.00	0.119	0.6
2006-07-27	24.4	4.4	0.50	0.101	1.1
2006-07-28	24.4	9.9	1.00	0.315	2.7
2006-07-29	26.1	2.5	0.00	-0.246	1.1
2006-07-30	26.7	4.4	0.00	0.063	0.0
2006-07-31	26.7	2.7	0.00	-0.112	0.9
2006-08-01	27.2	2.0	0.00	-0.023	0.5
2006-08-02	27.2	3.2	0.00	0.026	0.0
2006-08-03	26.7	7.8	0.50	0.296	0.0
2006-08-04	25.6	40.6	3.50	-0.122	0.2
2006-08-05	26.1	40.1	3.00	-0.003	0.1
2006-08-06	25.6	13.3	0.67	-0.141	0.0
2006-08-07	26.7	4.5	0.00	-0.125	0.1
2006-08-14	25.0	6.8	0.00	-0.157	0.0
2006-08-15	26.1	28.8	2.50	0.101	0.3
2006-08-16	25.0	6.9	0.50	0.020	0.1
2006-08-17	26.1	5.3	0.33	0.000	0.0
2006-08-18	25.6	6.8	0.00	-0.072	0.0
2006-08-19	24.4	3.5	0.00	0.062	0.4
2006-08-20	23.9	29.8	2.50	0.013	0.5
2006-08-21	23.3	32.9	1.50	-0.105	0.1
2006-08-22	25.0	9.0	0.00	-0.009	0.0
2006-08-23	25.0	6.2	0.50	0.085	0.0
2006-08-24	24.4	5.9	0.00	0.006	0.0
2006-08-28	24.4	32.2	1.50	0.069	0.4
2006-08-29	24.4	258.0	5.00	0.640	2.6
2006-08-30	23.3	162.8	2.50	-0.558	1.4
2006-08-31	23.3	83.0	4.00	0.229	0.0

Could you classify the following 56 rows using the Ecoli_binary label, where 1 indicates non-

Date	Lake_Temp_C	Lake_Turb_NTRU	WaveHt_Ft	LL_PreDay	AirportRain48W_in
2005-05-25	13.3	58.0	1.00	-0.099	0.0
2005-05-26	14.4	11.5	1.00	-0.256	0.0
2005-05-31	17.8	3.2	0.00	0.027	0.1
2005-06-01	17.8	3.5	0.50	0.045	0.0
2005-06-02	17.2	8.4	1.00	-0.062	0.0
2005-06-06	18.3	7.1	0.33	0.184	0.1
2005-06-07	20.0	9.3	0.50	-0.328	0.1
2005-06-08	20.0	3.0	0.00	0.108	0.0
2005-06-09	23.3	1.9	0.00	0.026	0.0
2005-06-13	24.4	2.9	0.00	0.079	0.1
2005-06-14	21.7	5.2	1.00	-0.069	0.3
2005-06-15	21.1	18.5	1.00	-0.148	1.1

2005-06-16	20.0	54.5	1.50	0.184	1.3
2005-06-20	22.2	32.5	0.50	-0.217	0.0
2005-06-21	21.7	26.0	0.50	-0.029	0.0
2005-06-22	22.8	23.0	0.50	0.029	0.2
2005-06-23	22.8	25.5	0.50	0.027	0.1
2005-06-27	24.4	9.8	0.00	-0.075	0.0
2005-06-28	23.9	4.2	0.00	0.049	0.0
2005-06-29	24.4	7.0	0.50	0.072	0.3
2005-06-30	24.4	3.6	0.00	-0.144	0.3
2005-07-05	24.4	6.2	1.00	0.069	0.0
2005-07-06	23.3	16.0	1.50	0.036	0.0
2005-07-07	24.4	12.0	1.00	-0.099	0.0
2005-07-11	24.4	3.8	0.50	0.004	0.0
2005-07-12	24.4	3.0	0.50	-0.073	0.0
2005-07-13	25.6	4.7	0.50	-0.009	0.0
2005-07-14	24.4	4.4	0.50	0.006	0.1
2005-07-18	25.0	4.4	0.50	-0.014	0.2
2005-07-19	24.4	4.3	0.50	0.050	0.4
2005-07-20	26.1	1.7	0.50	0.013	0.2
2005-07-21	23.9	6.1	0.50	-0.138	1.5
2005-07-25	25.6	8.3	0.50	-0.053	0.5
2005-07-26	25.6	2.3	0.50	-0.236	0.3
2005-07-27	24.4	62.5	5.00	0.857	2.9
2005-07-28	22.2	14.7	2.00	-0.584	1.9
2005-08-01	26.1	4.7	0.00	-0.003	0.0
2005-08-02	26.7	3.0	0.50	-0.053	0.0
2005-08-03	27.2	2.0	0.00	0.007	0.0
2005-08-04	26.1	1.6	0.50	-0.089	0.0
2005-08-08	26.7	1.7	1.00	-0.052	0.0
2005-08-09	26.7	1.5	0.00	-0.138	0.0
2005-08-10	26.1	1.9	0.00	0.023	0.0
2005-08-11	26.1	8.0	2.00	0.253	0.3
2005-08-15	25.6	13.5	2.00	0.062	0.3
2005-08-16	24.4	4.8	0.33	-0.095	0.0
2005-08-17	25.6	2.3	0.17	-0.118	0.0
2005-08-18	25.0	2.3	0.66	-0.040	0.0
2005-08-22	24.4	86.5	3.00	0.076	3.6
2005-08-23	24.4	28.5	4.00	0.023	0.1
2005-08-24	24.4	10.2	2.50	-0.050	0.1
2005-08-25	24.4	10.0	2.00	-0.032	0.0
2005-08-29	23.3	2.7	0.50	-0.282	1.2
2005-08-30	23.3	2.4	0.50	0.207	0.0
2005-09-01	22.8	52.0	2.33	-0.167	2.2
2005-09-06	21.7	21.5	1.00	-0.229	0.0

Please show the result in Python list format.