Organized Plasticity for Cost-Bounded Continual Adaptation

Anonymous Authors[1]

Abstract

Deployed AI systems must adapt to changing inputs, environments, user patterns, and operating constraints. Sustainable adaptation cannot route every deployment failure to a single dominant learner, since many failures are local, recurrent, and cost-sensitive. This paper proposes *organized plasticity*, a systems-level methodology that allocates durable adaptation across distinct loci of change: input stabilization, skill compilation, memory consolidation, and executive control. This organization enables systems to stabilize poor observations before deeper inference, compile repeated successes into cheaper routines, retain reusable knowledge without broad retraining, and adjust routing or update policies without perturbing task models. A compact feasibility evaluation in a dynamic-vision setting shows that input stabilization, skill compilation, and executive control improve the accuracy-cost tradeoff in different ways while keeping the dominant learner fixed. The results motivate evaluating continual adaptation by where durable change resides, what it costs, and what collateral effects it creates, rather than by final task performance alone.

1. Introduction

Deployed AI systems need continual adaptation, but the cost of adaptation depends not only on what changes, but also on where that change is retained. As inputs, environments, user patterns, and resource constraints shift after release, systems must correct behavior without repeatedly relying on larger contexts, memory expansion, teacher calls, model editing, or fine-tuning (Wang et al., 2024b; Packer et al., 2023; Gu et al., 2024; Gupta et al., 2024). When local failures are routed to one dominant corrective surface, triage becomes opaque, recurring cases remain expensive, and adaptation turns into a retrain-redeploy loop rather than cumulative competence (Sculley et al., 2015; Amershi et al., 2019; Kreuzberger et al., 2023).

The key question is therefore not only whether deployed systems can adapt, but where durable correction should reside so that future behavior improves with minimal cost and collateral risk. A degraded observation, a repeated low-confidence success, a missing memory, and a poor decision about when to escalate from a cheap local path to a higher-cost reasoning path may all appear as task errors, but they point to different corrective mechanisms. Some require improved sensing or context formation, while others require a reusable routine, a memory update, an escalation-policy adjustment, or a broader model update.

We therefore frame post-deployment adaptation as a problem of localizing plasticity. Durable correction should occur at the cheapest corrective surface that can address the failure under latency, compute, memory, communication, and reliability constraints. Organized plasticity makes the location of durable change explicit. The proposed scaffold separates four loci of change: input stabilization for improving inputs or context, skill compilation for making recurring cases cheaper, memory consolidation for retaining high-value experience, and executive control for routing, escalation, abstention, and update gating (McClelland et al., 1995; Graybiel, 1998; Miller & Cohen, 2001; Yang et al., 2016). Its role is to make post-deployment adaptation inspectable, comparable, and testable under matched resource budgets. A compact case study instantiates this scaffold, illustrating how degraded observations, recurring cases, and unnecessary escalation can be handled by different corrective loci without updating the dominant learner.

The main contributions are:

- **Localization principle** for cost-bounded continual adaptation, framing post-deployment change by where it resides.
- **Four-locus scaffold** that separates adaptation by trigger, persistence, budget, and collateral risk.
- **Feasibility case study** in a dynamic-vision setting that illustrates how different failure sources can be handled by different corrective loci, reported through accuracy and higher-cost cloud-call rate.

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the ICML 2026 Workshop "Continual Adaptation at Scale: Towards Sustainable AI". Do not distribute.

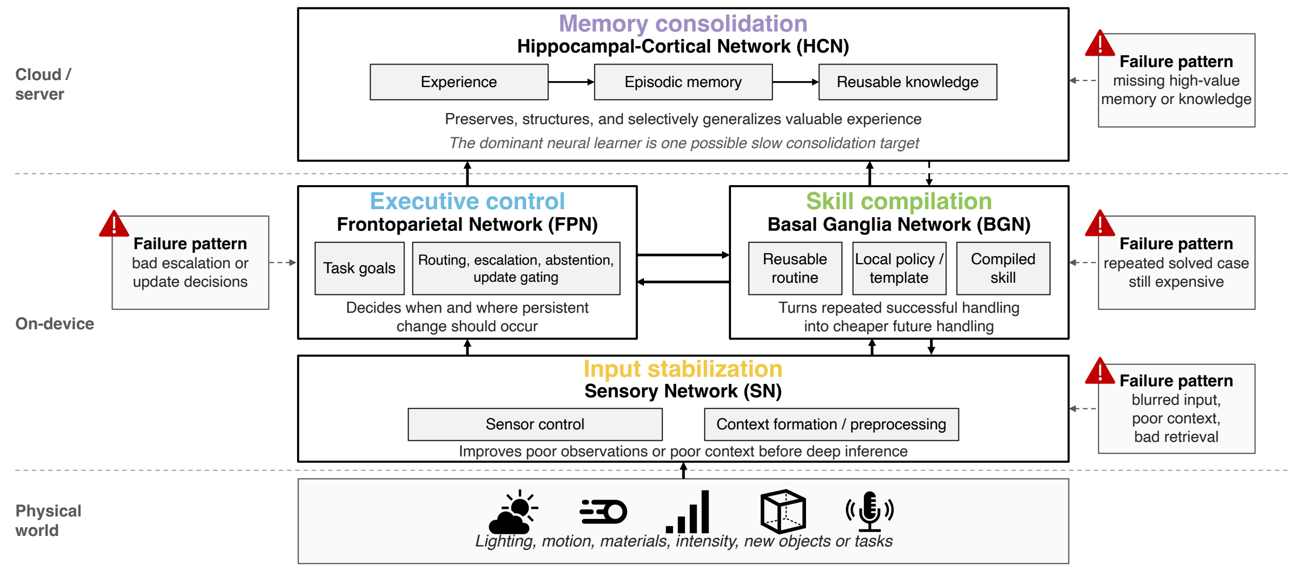

Figure 1. **Four loci organize persistent change in cost-bounded AI systems.** Functional roles map to four loci of plasticity, SN for input stabilization, BGN for skill compilation, HCN for memory consolidation, and FPN for executive control. Each locus supports a distinct kind of durable change, occupies a deployment tier from local to cloud, and is triggered by a characteristic failure pattern. Arrows summarize the principal dependencies: SN conditions the signal consumed by downstream loci, FPN gates escalation from BGN to HCN as well as broader update decisions, and HCN handling of recurring cases can be compiled into BGN routines.

2. Organized Plasticity

A locus of plasticity is any system element that can retain experience-dependent change and alter future behavior, such as a policy, memory store, routine, threshold, controller, or learned module. Loci are distinguished not by implementation type, but by the kind of persistent change they support and by the trigger, persistence, budget, and collateral effect associated with that change. Organized plasticity localizes post-deployment adaptation by asking where durable correction should reside. It differs from selective computation, which asks where inference should go on the current episode (Fedus et al., 2022; Snell et al., 2025). The focus here is instead where persistent change should accumulate across episodes.

2.1. Four Functional Loci

The four-locus scaffold makes this localization concrete. It uses the brain as a functional reference, not a blueprint. Biological intelligence separates adaptation into four burdens, namely maintaining usable observations before deep inference, turning repeated success into cheap routine, preserving and consolidating experience as reusable knowledge, and gating retrieval, escalation, inhibition, and learning itself (McClelland et al., 1995; Graybiel, 1998; 2008; Miller & Cohen, 2001; Yang et al., 2016). These burdens define four motifs: Sensory Network (SN) for input stabilization, Basal Ganglia Network (BGN) for skill compilation, Hippocampal-Cortical Network (HCN) for memory consolidation, and Frontoparietal Network (FPN) for executive control. Figure 1 summarizes this mapping.

Input stabilization (SN) addresses failures caused by degraded observation or context quality. In physical systems, it adjusts exposure, gain, viewpoint, or active sensing (Yang et al., 2016; Baek et al., 2025; Choi et al., 2026); in language and agent systems, it reformulates queries, filters retrieval, or summarizes state (Packer et al., 2023).

Skill compilation (BGN) addresses repeated cases that remain unnecessarily expensive. Converting a recurring pattern into a routine, template, or local controller prevents deployment from becoming repeated rediscovery. The idea connects biological habit formation (Graybiel, 1998; 2008) with agent systems that reuse learned routines or skills (Wang et al., 2024a).

Memory consolidation (HCN) addresses sparse or high-value experience that should affect future episodes. Some experiences remain episodic, some are abstracted into reusable knowledge, and some eventually seed broader model updates (McClelland et al., 1995; Varga et al., 2024; Dorovatas et al., 2026). The dominant neural learner is one possible slow consolidation target inside HCN, not the default destination for every correction.

Executive control (FPN) decides whether, where, and how adaptation should occur, through routing, escalation, abstention, update gates, and safety vetoes (Miller & Cohen, 2001). This role is necessary because persistent change requires a control policy for when to act locally, when to escalate, and when to allow or block updates.

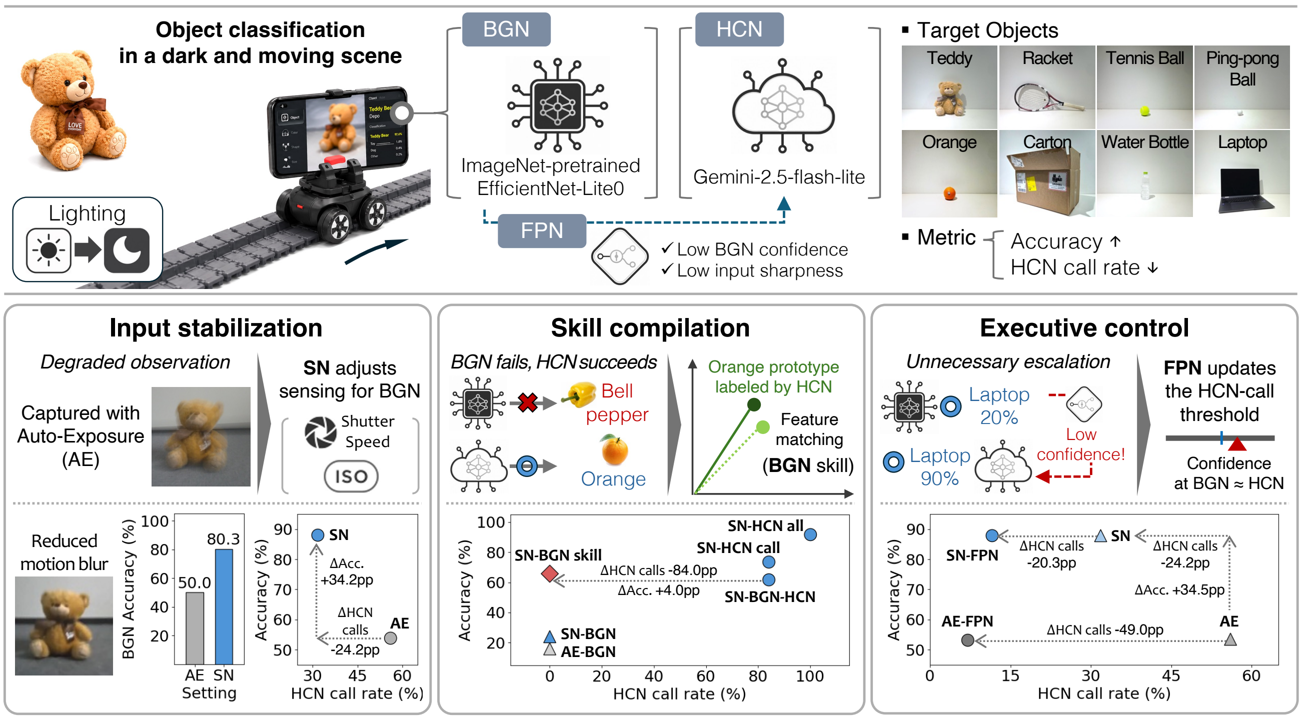

Figure 2. **Localized adaptation in a dynamic-vision case study.** Eight physical target objects are classified under a dark-and-motion shift, where an ImageNet-pretrained EfficientNet-Lite0 BGN classifier faces smartphone images captured from a moving platform rather than stable static inputs. We report accuracy and HCN call rate, the latter measuring reliance on the higher-cost cloud-side recognizer, Gemini-2.5-flash-lite. The three panels show localized corrections: SN improves acquisition, BGN compiles a recurring HCN-solvable case into a feature-template routine, and FPN adjusts the BGN-confidence threshold to reduce unnecessary HCN calls.

2.2. Locus Composition and Auditable Correction

The four loci compose rather than act in isolation. Input stabilization at the SN locus improves the signal consumed by BGN and HCN, preempting downstream corrections that degraded input would otherwise trigger. Skill compilation at the BGN locus absorbs recurring cases that previously required HCN, redirecting future episodes away from the cloud and reducing the rate at which FPN must arbitrate escalation. HCN-side consolidation can in turn seed BGN templates, allowing a case first handled by deeper reasoning to be compiled into a local routine after repeated success. FPN coordinates these flows by controlling local acceptance, escalation, abstention, and update gating.

This composition makes correction auditable because each improvement can be attributed to a locus with a characteristic cost and collateral risk. The same task accuracy may be reached through different corrective loci, but the resulting adaptation paths are not equivalent. A monolithic baseline may recover performance through repeated prompting, teacher calls, memory expansion, model editing, or fine-tuning, but it often obscures where correction resides and what it costs. An organized-plasticity system instead asks which corrective locus recovers performance most cheaply under constraints.

3. Feasibility Evaluation

3.1. Objective and Deployment Setting

This evaluation examines whether localized adaptation can improve the accuracy-cost tradeoff under deployment shift. The goal is not to prove that the four-locus scaffold is optimal, but to illustrate how different sources of mismatch can be handled by different corrective loci rather than by a single dominant learner. Figure 2 summarizes the three localized adaptation experiments.

The task is object classification in a dynamic-vision setting, building on the smartphone-based testbed introduced by Choi et al. (2026). A smartphone-mounted camera observes tabletop objects from a moving small vehicle after shifting from stable visual conditions to dark, moving conditions, where reduced illumination and motion blur degrade recognition. The local path uses an ImageNet-pretrained (Deng et al., 2009) EfficientNet-Lite0 (Tan & Le, 2019) as the BGN classifier, while a higher-cost cloud-side recognizer, Gemini-2.5-flash-lite (Comanici et al., 2025), serves as HCN when escalation is needed. We report accuracy as task performance and HCN call rate as a proxy for reliance on the higher-cost reasoning path. These shifts are simple by design, but they capture a common deployment pattern: the same classification failure may admit different lower-overhead corrections, including better sensing, local routine formation, or a revised escalation policy.

3.2. Localized Adaptation Experiments

We organize the feasibility evaluation around three failure modes and their localized corrections. Adaptation is restricted to SN, BGN, and FPN, while HCN is kept fixed because updating the dominant learner is the costly path.

Input stabilization (SN). The first failure mode is degraded observation under dark and moving conditions. Default auto-exposure (`AE`) often produces dim or motion-blurred frames, reducing local recognition and increasing escalation. Replacing `AE` with SN-controlled acquisition (`SN`) improves the input before inference, allowing more cases to resolve locally. Local BGN accuracy rises from 50.0% to 80.3%, end-to-end accuracy with HCN fallback rises from 53.8% to 88.0%, and HCN call rate drops from 56.0% to 31.8%. This illustrates the composition described in Section 2.2: input stabilization at the SN locus preempts downstream load that BGN and HCN would otherwise absorb.

Skill compilation (BGN). The second failure mode is a recurring local failure that HCN can solve. For the orange class, `AE-BGN` achieves 16.0% accuracy, and SN-controlled acquisition improves `SN-BGN` only modestly to 24.0%, indicating that input stabilization alone does not resolve the case. In contrast, sending SN-stabilized inputs directly to HCN yields a 92.0% accuracy upper bound (`SN-HCN all`). The selective escalation path, `SN-BGN-HCN`, escalates 84.0% of evaluation episodes to HCN and reaches 62.0% end-to-end accuracy. On the escalated subset, HCN is correct in 73.8% of cases (`SN-HCN call`), but the remaining local decisions are still dominated by BGN errors. We therefore compile the HCN-labeled orange representation into a BGN-side feature template and classify future cases by cosine similarity. This `SN-BGN skill` path eliminates HCN calls for the compiled decisions and improves local accuracy to 66.0%, showing how repeated cloud-handled success can become a cheaper local routine.

Executive control (FPN). The third failure mode is unnecessary escalation on locally resolvable cases. The laptop case in Figure 2 illustrates this pattern, while the results average over all eight target object classes. Here, `AE` and `SN` denote default auto-exposure and SN-controlled acquisition with the original escalation rule, while `AE-FPN` and `SN-FPN` add FPN-tuned gating. FPN adjusts the BGN-confidence threshold to the range where HCN escalation no longer changes the prediction because BGN and HCN agree. Under AE, FPN reduces HCN call rate from 56.0% to 7.0% with accuracy unchanged within noise, from 53.8% to 53.2%. Combined with SN, FPN reduces HCN call rate from 31.8% to 11.5% while preserving 88.0% accuracy. Thus, executive control removes unnecessary higher-cost reasoning calls without changing the task model.

3.3. Analysis and Implications

The central feasibility result is that failures appearing as the same classification error admit distinct lower-cost corrections. Input stabilization at the SN locus improves what the system observes, raising the accuracy ceiling while reducing downstream HCN calls. Skill compilation at BGN turns a recurring HCN-solvable case into a local routine, shifting the cost frontier as the case repeats. Executive control at FPN reduces unnecessary escalation by revising the BGN-confidence threshold, without changing the task model. Together, these results show that useful durable correction need not occur in the dominant learner.

The remaining question is how to identify the cheapest locus for durable change. Automatic locus triage could use failure signatures to select the appropriate corrective surface, while cross-locus promotion rules would determine when memory becomes routine or when repeated input failures should trigger control-policy changes. These mechanisms are needed to make organized plasticity operate automatically under deployment shift.

4. Conclusion

Cost-bounded continual adaptation is not only a scaling problem, but also a problem of localizing durable change. Rather than treating every post-deployment failure as a signal to update one dominant learner, organized plasticity distributes durable correction across input stabilization, skill compilation, memory consolidation, and executive control. In our feasibility evaluation, input stabilization recovered much of the accuracy gap while reducing escalation, executive control cut cloud calls by roughly three to eight times without measurable accuracy loss, and skill compilation eliminated cloud calls for a recurring case with only a small accuracy concession. HCN-side memory consolidation remains outside this evaluation, but the scaffold clarifies when such higher-cost adaptation is warranted: after cheaper loci fail to correct the recurring mismatch.

The broader implication is methodological. The same task accuracy can be reached through corrections with very different costs, and final task accuracy alone cannot reveal which correction was used or whether a cheaper one existed. We therefore suggest that adaptive-AI evaluations report where durable change resides, what triggered it, how long it persists, what it costs, and what prior competence it preserves or damages. Making these quantities explicit allows adaptive systems to be compared under matched resource constraints and judged by whether they become cheaper, safer, and more reliable across repeated deployment.

References

Amershi, S., Begel, A., Bird, C., DeLine, R., Gall, H., Kamar, E., Nagappan, N., Nushi, B., and Zimmermann, T. Software engineering for machine learning: A case study. In *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*, pp. 291–300, 2019. doi: 10.1109/ICSE-SEIP.2019.00042.

Baek, E., Park, K., Ko, J., Oh, M., Gong, T., and Kim, H.-S. Position: AI should sense better, not just scale bigger: Adaptive sensing as a paradigm shift. In *The Thirty-Ninth Annual Conference on Neural Information Processing Systems Position Paper Track*, 2025. URL https://openreview.net/forum?id=xnNHXepQ9h.

Choi, Y. R., Park, S., and Kim, H.-S. [emerging ideas] artificial tripartite intelligence: A bio-inspired, sensor-first architecture for physical ai. *arXiv preprint arXiv:2604.13959*, 2026.

Comanici, G., Bieber, E., Schaekermann, M., Pasupat, I., Sachdeva, N., Dhillon, I., Blistein, M., Ram, O., Zhang, D., Rosen, E., et al. Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. *arXiv preprint arXiv:2507.06261*, 2025.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, 2009. doi: 10.1109/CVPR.2009.5206848.

Dorovatas, V., Schwerin, M., Bagdanov, A. D., Caccia, L., Carta, A., Charlin, L., Hammer, B., Hayes, T. L., Hess, T., Kanan, C., et al. Modular memory is the key to continual learning agents. *arXiv preprint arXiv:2603.01761*, 2026.

Fedus, W., Zoph, B., and Shazeer, N. Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity. *Journal of Machine Learning Research*, 23(120):1–39, 2022. URL http://jmlr.org/papers/v23/21-0998.html.

Graybiel, A. M. The basal ganglia and chunking of action repertoires. *Neurobiology of Learning and Memory*, 70(1):119–136, 1998. ISSN 1074-7427. doi: https://doi.org/10.1006/nlme.1998.3843. URL https://www.sciencedirect.com/science/article/pii/S1074742798938436.

Graybiel, A. M. Habits, rituals, and the evaluative brain. *Annu. Rev. Neurosci.*, 31(1):359–387, 2008.

Gu, J.-C., Xu, H.-X., Ma, J.-Y., Lu, P., Ling, Z.-H., Chang, K.-W., and Peng, N. Model editing harms general abilities of large language models: Regularization to the rescue. In Al-Onaizan, Y., Bansal, M., and Chen, Y.-N. (eds.), *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pp. 16801–16819, Miami, Florida, USA, November 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.emnlp-main.934. URL https://aclanthology.org/2024.emnlp-main.934/.

Gupta, A., Rao, A., and Anumanchipalli, G. Model editing at scale leads to gradual and catastrophic forgetting. In Ku, L.-W., Martins, A., and Srikumar, V. (eds.), *Findings of the Association for Computational Linguistics: ACL 2024*, pp. 15202–15232, Bangkok, Thailand, August 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.findings-acl.902. URL https://aclanthology.org/2024.findings-acl.902/.

Kreuzberger, D., Kühl, N., and Hirschl, S. Machine learning operations (mlops): Overview, definition, and architecture. *IEEE Access*, 11:31866–31879, 2023. doi: 10.1109/ACCESS.2023.3262138.

McClelland, J. L., McNaughton, B. L., and O'Reilly, R. C. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological review*, 102 3:419–457, 1995. URL https://api.semanticscholar.org/CorpusID:2832081.

Miller, E. K. and Cohen, J. D. An integrative theory of prefrontal cortex function. *Annual review of neuroscience*, 24(1):167–202, 2001.

Packer, C., Wooders, S., Lin, K., Fang, V., Patil, S. G., Stoica, I., and Gonzalez, J. E. MemGPT: towards LLMs as operating systems. *arXiv preprint arXiv:2310.08560*, 2023.

Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.-F., and Dennison, D. Hidden technical debt in machine learning systems. In Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015. URL https://proceedings.neurips.cc/paper_files/paper/2015/file/86df7dcfd896fcaf2674f757a2463eba-Paper.pdf.

Snell, C. V., Lee, J., Xu, K., and Kumar, A. Scaling LLM test-time compute optimally can be more effective than scaling parameters for reasoning. In *The Thirteenth International Conference on Learning Representations*,

2025. URL https://openreview.net/forum?id=4FWAwZtd2n.

Tan, M. and Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pp. 6105–6114. PMLR, 2019.

Varga, N. L., Morton, N. W., and Preston, A. R. Schema, inference, and memory. In Kahana, M. J. and Wagner, A. D. (eds.), *The Oxford Handbook of Human Memory, Two Volume Pack: Foundations and Applications*. Oxford University Press, 06 2024. ISBN 9780190917982. doi: 10.1093/oxfordhb/9780190917982.013.47. URL https://doi.org/10.1093/oxfordhb/9780190917982.013.47.

Wang, G., Xie, Y., Jiang, Y., Mandlekar, A., Xiao, C., Zhu, Y., Fan, L., and Anandkumar, A. Voyager: An open-ended embodied agent with large language models. *Transactions on Machine Learning Research*, 2024a. ISSN 2835-8856. URL https://openreview.net/forum?id=ehfRiF0R3a.

Wang, L., Zhang, X., Su, H., and Zhu, J. A comprehensive survey of continual learning: Theory, method and application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(8):5362–5383, 2024b. doi: 10.1109/TPAMI.2024.3367329.

Yang, S. C.-H., Wolpert, D. M., and Lengyel, M. Theoretical perspectives on active sensing. *Current Opinion in Behavioral Sciences*, 11:100–108, 2016. ISSN 2352-1546. doi: https://doi.org/10.1016/j.cobeha.2016.06.009. URL https://www.sciencedirect.com/science/article/pii/S2352154616301255.