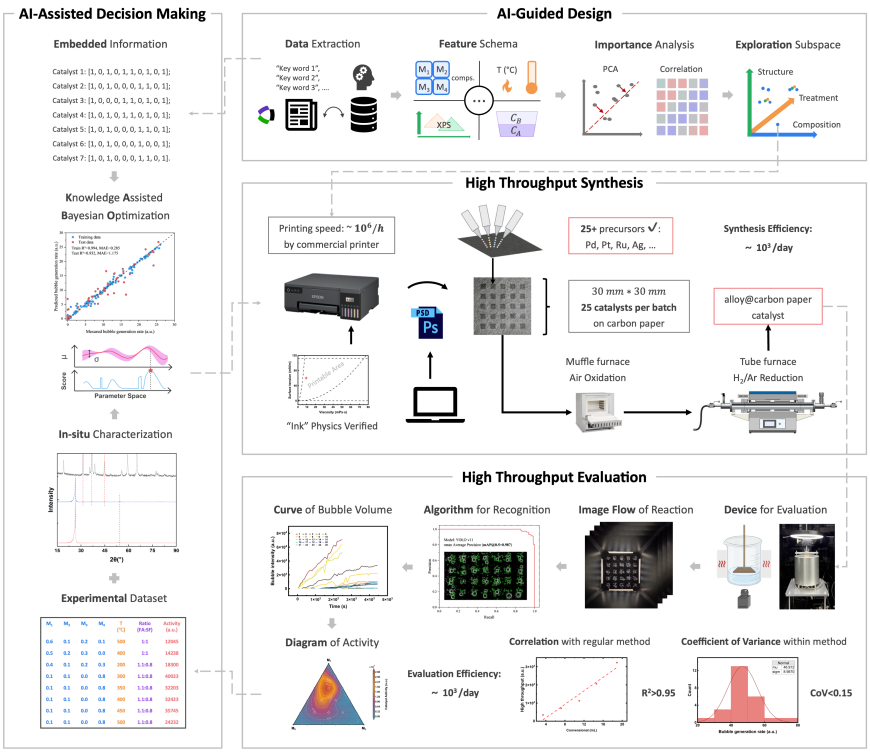
A Data-driven Closed-looped High-throughput Platform for Thermocatalyst Discovery

Zhengzuo Liu[a], Binhang Yan[a*]

[a]*Department of Chemical Engineering, Tsinghua University, Beijing 100084, China*

*Corresponding author.

E-mail address: binhangyan@tsinghua.edu.cn.



The transition toward a carbon-neutral energy landscape relies on the efficient utilization of hydrogen as a clean energy vector, yet widespread deployment is hindered by safety and logistical challenges in storage and transportation. Liquid Organic Hydrogen Carriers (LOHCs) provide a practical route by leveraging existing liquid-fuel infrastructure. Among them, formic acid (FA) is particularly attractive due to its high volumetric hydrogen capacity (53 g H_2 L^{-1}), low toxicity, and the ability to release hydrogen under mild conditions via formic acid dehydrogenation (FA-DH). Palladium (Pd)-based catalysts are among the most active for FA-DH, but mono-metallic systems

are nearing intrinsic performance ceilings, while the cost of noble metals motivates multi-metallic alloy catalysts that maximize atomic efficiency and catalytic turnover.

Alloying can modulate the electronic and geometric structures of active sites to circumvent scaling relations; however, extending from binary to higher-order alloys triggers a combinatorial explosion ("curse of dimensionality"). Traditional intuition-driven trial-and-error is inefficient and often biased toward familiar high-performing regions (e.g., Pd–Ag or Pd–Au), leaving vast "dark" areas of chemical space unexplored. To systematically access these non-intuitive regions, a shift from low-throughput, experience-based discovery to high-throughput, data-driven exploration is required.

Here we report a fully integrated self-driving laboratory (SDL) for closed-loop discovery of FA-DH catalysts. On the hardware side, we developed an array-based synthesis workflow using microfluidic inkjet printing, coupled with computer-vision bubble recognition for parallel activity evaluation. The platform achieves a screening throughput exceeding 10^3 samples per day—over three orders of magnitude faster than conventional benchtop workflows. Importantly, the micro-synthesis is designed for extreme atomic economy: by dispensing picolitre-scale droplets under a "constant total consumption" strategy, the total noble-metal usage required to synthesize 1,000 distinct samples is comparable to that of a single conventional batch experiment, making comprehensive traversal of expensive alloy spaces economically viable.

High-throughput experimentation alone remains inefficient without intelligent navigation. We therefore integrated a two-layer AI decision engine. First, to define a scientifically grounded exploration boundary, a large language model (LLM) performed semantic mining over ~250 relevant publications and proposed a 6-element compositional space that balances exploitation of known active elements with exploration of potential promoters. Second, we benchmarked multiple navigation strategies under a data-rich regime, including (i) statistical design of experiment (DOE) driven screening, (ii) conventional Bayesian optimization (BO), (iii) a literature-informed AI agent enhanced BO. These comparisons indicate that incorporating knowledge priors can improve sample efficiency and robustness against premature convergence while preserving exploration into non-intuitive regions.

By synergizing high-throughput hardware with knowledge-driven AI, we identified a previously unreported catalyst formulation exhibiting top-tier performance relative to reported values. Notably, this competitiveness is achieved with a deliberately simple, scalable preparation that does not rely on ultrafine nanostructuring or elaborate supports, highlighting the power of composition-driven synergy. Overall, this work establishes a generalizable, cost-effective SDL paradigm for accelerating discovery in complex catalytic materials beyond human bias-limited search.