The Matrix Has You!

How Generative AI Is Designing the Molecules We Can't Yet Imagine

From active learning to generative chemistry: the next frontier for PFAS-free DWR

Vasav Sahni, PhD
Polymer scientist specializing in performance textile finishes and PFAS-free systems.

CONTENTS

Executive summary

- The regulatory clock is already ticking: as of 2025–2026, 18+ U.S. states have enacted PFAS textile bans, with California and New York bans live since January 2025 and EU restrictions on PFAS in textiles taking effect January 2026. Outdoor apparel for severe wet conditions faces its own deadline in 2028. The compliance window is narrowing.
- Paper #4 (Moneyball for Materials) established that AI is most valuable when it makes rigor scalable; prioritizing experiments, predicting robustness, and capturing institutional memory. This paper takes the next step: generative AI doesn't just optimize within a known chemical space. It proposes candidate molecules that chemists haven't considered yet.
- The shift from predictive to generative is the step-change. Bayesian optimization and supervised ML explore the space you've already mapped. Generative models: diffusion models, graph neural networks, GFlowNets etc. design novel molecular structures from scratch, constrained by desired properties: water repellency, wash durability, low toxicity, no persistent fluorine.
- IBM Research's MatGFN-PFAS and academic work using MolHGT+ (a hybrid graph neural network) demonstrate that AI can now propose fluorine-free or reduced-fluorine alternatives while explicitly optimizing for low bioaccumulation and low hepatotoxicity, not just functional performance.
- The synthesizability gap is closing. New frameworks like SynFormer (PNAS, 2025) ensure that every AI-generated molecule comes with a viable synthetic pathway thus bridging the gap between in silico design and what can be made in a lab.
- The practical playbook: combine the active learning loop from Paper #4 with generative pre-screening. Use generative AI to propose novel PFAS-free DWR candidates -> filter for synthesizability and toxicity -> enter the active learning loop for formulation and process optimization -> validate at mill scale.

1.

Mandatory, Not Optional

In Paper #4 we framed the PFAS-free DWR challenge as a multi-variable optimization problem: finish chemistry, substrate variability, process windows, and durability all working together. The AI playbook we described - active learning, Bayesian optimization, minimum viable datasets - was about getting smarter inside a defined formulation space.

The regulatory landscape is tightening rapidly in ways that changes the scope of the problem itself. We are no longer in a world where PFAS-free DWR is a competitive advantage. It is rapidly becoming a compliance requirement.

Current Regulatory Landscape

STATE	REGULATION	WHEN
CALIFORNIA	Ban on intentionally added PFAS in all textile articles (AB-1817)	Jan 2025 (100 ppm); Jan 2027 (50 ppm)
NEW YORK	Ban on PFAS in all new apparel; outdoor/severe-wet exemption ends 2028	Jan 2025
EU (REACH)	Ban on PFAS in textiles, footwear, waterproofing agents for consumers	Jan 2026
MINNESOTA	PFAS ban in fabric treatments, textile furnishings; reporting due Jul 2026	Jan 2025 / Jul 2026
NEW MEXICO	Full product PFAS ban including textiles (unless unavoidable use)	Jan 2028 —> 2032
18+ U.S. STATES	~200 PFAS bills introduced annually; patchwork of bans, labeling, reporting	2026–2032 wave
FEDERAL (EPA/TSCA)	PFAS reporting rule; federal compliance moderated but state laws accelerating	Ongoing; 2026 reporting deadlines

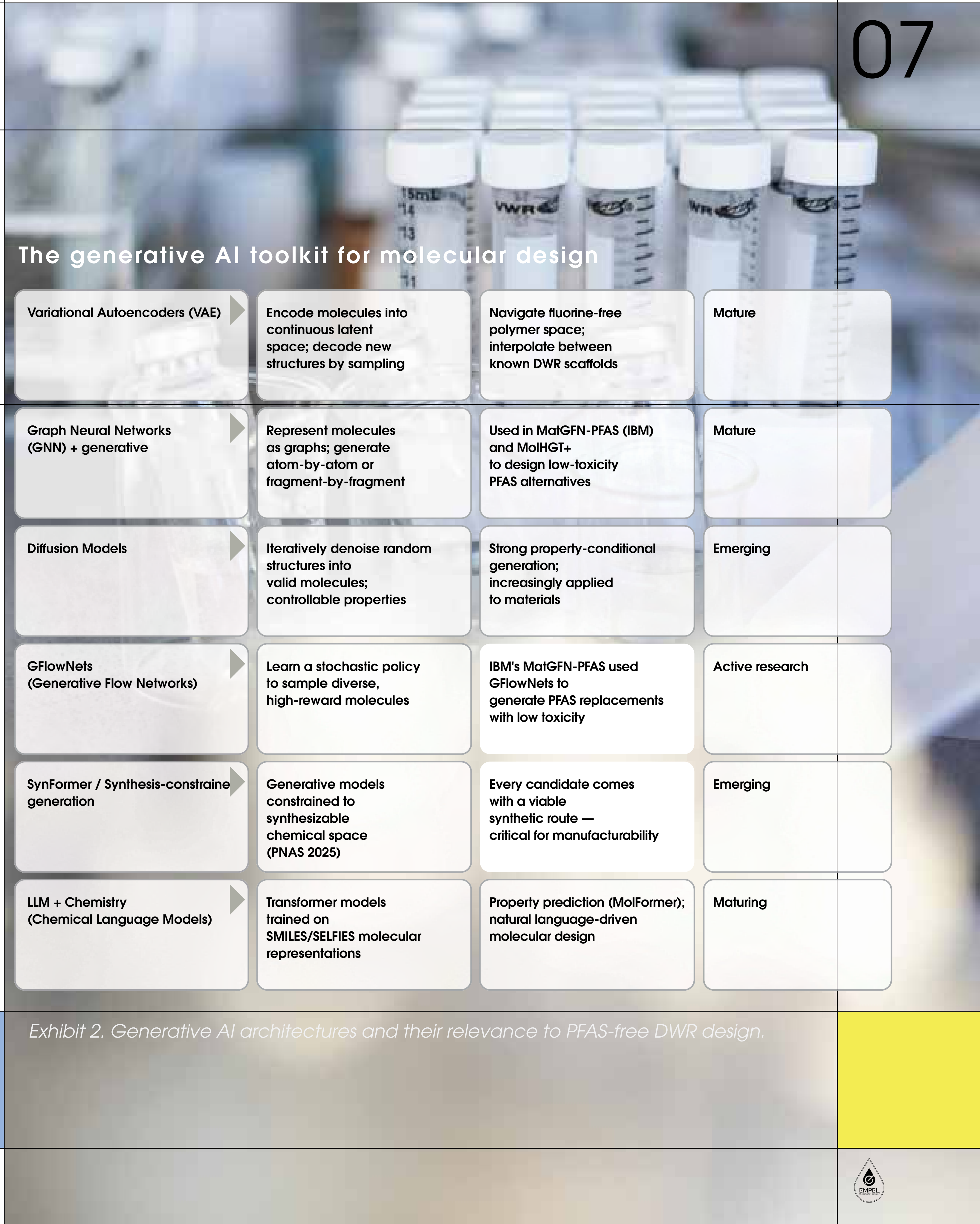
Exhibit 1. Selected PFAS restrictions impacting textile DWR as of early 2026. Sources: MultiState (2026), Trimco Group (2025), Morgan Lewis (2024–2026), Arnold & Porter (2025).

2.

From Optimization to Invention: The Generative Step-Change

Paper #4 described AI as a better scout; one that tells you which experiments to run next. Generative AI is a different kind of tool. It doesn't navigate the space you already know. It proposes addresses in neighborhoods that don't exist on your current map.

This matters for PFAS-free DWR because the challenge is not just reformulation. It is molecular invention: finding polymers and surface-active agents that deliver comparable water-repellency and wash durability without the C–F bonds that make PFAS both effective and persistent. The candidates that satisfy all constraints simultaneously - performance, durability, processability, low toxicity, regulatory compliance, and cost - may not exist in any existing ingredient library. They may need to be designed.



The generative AI toolkit for molecular design

Variational Autoencoders (VAE)	Encode molecules into continuous latent space; decode new structures by sampling	Navigate fluorine-free polymer space; interpolate between known DWR scaffolds	Mature
Graph Neural Networks (GNN) + generative	Represent molecules as graphs; generate atom-by-atom or fragment-by-fragment	Used in MatGFN-PFAS (IBM) and MolHGT+ to design low-toxicity PFAS alternatives	Mature
Diffusion Models	Iteratively denoise random structures into valid molecules; controllable properties	Strong property-conditional generation; increasingly applied to materials	Emerging
GFlowNets (Generative Flow Networks)	Learn a stochastic policy to sample diverse, high-reward molecules	IBM's MatGFN-PFAS used GFlowNets to generate PFAS replacements with low toxicity	Active research
SynFormer / Synthesis-constraine generation	Generative models constrained to synthesizable chemical space (PNAS 2025)	Every candidate comes with a viable synthetic route — critical for manufacturability	Emerging
LLM + Chemistry (Chemical Language Models)	Transformer models trained on SMILES/SELFIES molecular representations	Property prediction (MolFormer); natural language-driven molecular design	Maturing

Exhibit 2. Generative AI architectures and their relevance to PFAS-free DWR design.

IBM Research MatGFN-PFAS
(2024)

MolHGT+ for Environmentally Friendly PFAS Design
(ScienceDirect, 2024)

SynFormer: Synthesizable Generative Design
(PNAS, 2025)

Self-Driving Labs and Closed-Loop Validation
(Nature, 2023; Chemical Reviews, 2024)

Nobel Prize validation
(2024)

3.

The Proof Points: What Generative AI Has Already Done for PFAS and Materials

IBM Research MatGFN-PFAS (2024):
IBM deployed GFlowNets combined with the MolFormer chemical language model to generate non-toxic alternatives to PFAS superacids. The system used two design strategies: Tversky similarity (generating molecules structurally like PFAS but with reduced toxicity) and direct property optimization (molecules with target functionality but no fluorinated backbone). This is the closest existing proof point to PFAS-free DWR molecular design.

MolHGT+ for Environmentally Friendly PFAS Design (ScienceDirect, 2024):
A hybrid deep learning architecture combining heterogeneous graph neural networks with transformer-like attention was used to predict surface tension, bioaccumulation, and hepatotoxicity. Virtual screening identified that siloxane groups and betaine fragments can preserve low surface tension while substantially reducing both bioaccumulation and toxicity — directly relevant to water-repellent finishes.

SynFormer: Synthesizable Generative Design (PNAS, 2025):
A generative framework ensuring every proposed molecule has a viable synthetic pathway. The persistent failure mode for generative AI in chemistry has been proposing molecules that cannot be made. SynFormer closes this gap, making AI-designed candidates actionable.

Self-Driving Labs and Closed-Loop Validation (Nature, 2023; Chemical Reviews, 2024):
Autonomous laboratory platforms now combine AI-proposed synthesis routes with robotic execution and property measurement. The generative design —> synthesis —> testing loop is no longer purely theoretical — it is being operationalized, particularly in academic and pharma settings.

Nobel Prize validation (2024):
The 2024 Nobel Prize in Chemistry was awarded for breakthroughs in protein structure prediction and AI-designed proteins (AlphaFold; David Baker's work on de novo protein design). The signal: AI-driven molecular design has crossed from research curiosity to validated scientific methodology.

4.

The Integrated Playbook: Generative Design Meets the Active Learning Loop

Paper #4 described a five-step closed learning loop for PFAS-free DWR: define success gates -> instrument experiments -> use DOE then Bayesian optimization -> train for robustness -> make tradeoffs explicit. Generative AI adds a prior stage to this loop - one that expands the candidate set before you even begin formulation work.

The extended loop

Stage 0 — Generative Pre-screening: Define target properties for PFAS-free DWR: contact angle, wash durability, low or no persistent organic fluorine, processability at standard mill conditions, cost constraints. Use a generative model (GNN, VAE, or GFlowNet) to propose a diverse candidate library filtered against synthesizability (SynFormer-style) and toxicity (MolHGT+-style screens).
Stage 1 — Minimum Viable Dataset + Structure: From Stage 0, select 10–20 synthesizable candidates spanning structural diversity. Synthesize and enter them into the MVD framework from Paper #4 (formulation, process, substrate, readiness/contamination descriptors, outcomes).
Stage 2 — DOE + Bayesian Optimization: Run the structured experiment loop from Paper #4. The active learning model now has better starting candidates because generative pre-screening narrowed the space intelligently rather than relying on existing ingredient catalogs.
Stage 3 — Robustness and Scale: Train for substrate variability and contamination challenge as before. Add a feedback loop: when a candidate fails for structural reasons (not just formulation), return that failure signal to the generative model to constrain the next generation of proposals.
Stage 4 — Synthesizability and IP Check: Before committing to a candidate, verify the synthetic route is feasible at commercial scale and conduct a freedom-to-operate assessment. AI-designed molecules may be novel — that is an asset (patentable) but also a risk (unknown regulatory history). Build this gate explicitly.

The key insight: generative AI and active learning are not competing approaches. Generative design answers the question 'what should we try?' at the molecular level. Active learning answers 'what experiment should we run next?' at the formulation level. Together, they compress the search space from both ends.

What Makes PFAS-Free DWR a Hard (and Interesting) Generative Design Problem

Not all molecular design problems are equally amenable to generative AI. PFAS-free DWR is harder than drug discovery in some ways, and more tractable in others.

Why it's harder

- Multi-scale performance: a molecule must perform at the fiber surface, survive formulation chemistry, bond under mill curing conditions, and resist mechanical and hydrolytic degradation through 50+ wash cycles. Drug targets have a receptor binding site; DWR has a complex multi-condition test regime.
- Process coupling: the 'right' molecule depends on the substrate, bath chemistry, add-on level, cure temperature, and downstream finishing. Generative models trained on molecular structure alone will miss this coupling unless process context is embedded as conditioning variables.
- Sparse training data: drug discovery benefits from large public datasets (ChEMBL, PubChem). DWR performance data is proprietary, scattered, and often underdocumented. Small-data generative strategies are essential.

Why it's more tractable

- The property space is better-defined: contact angle, spray rating, durability after wash, and oil repellency are quantifiable. Generative models can be explicitly conditioned on these targets, unlike complex biological outcomes.
- Toxicity filters are increasingly reliable: PFAS-specific toxicity models (bioaccumulation, hepatotoxicity, environmental persistence) are now accurate enough to use as hard filters during generation, not just post-hoc screening.
- The competitive moat is real: companies that develop proprietary generative pipelines for PFAS-free DWR chemistry will build libraries of novel, patentable candidates that cannot be replicated by teams using traditional trial-and-error approaches.

5.

6.

The Honest Constraints: What Generative AI Still Cannot Do

It cannot replace wet chemistry expertise.
Generative models propose structures. A materials chemist must evaluate whether the proposed structure is compatible with textile processing, emulsifiable at use concentrations, and stable under curing conditions. AI and chemist are collaborators, not substitutes.

Training data quality is still the bottleneck.
A generative model trained on undocumented historical data will reproduce the biases and blind spots of that data. The MVD from Paper #4 is not optional; it is the foundation that makes generative AI useful rather than misleading.

Novel molecules need novel regulatory pathways.
A truly new molecular structure has no regulatory history. This is a feature (patent protection) but also a cost (toxicology studies, registration). Build the regulatory timeline into your AI roadmap from the start.

Synthesizability at commercial scale is a separate problem.
SynFormer-style constraints help, but academic synthesizability and commercial-scale polymer synthesis economics are different problems. The lab route and the plant route are not the same.

Multi-objective optimization is still hard.
Maximizing water repellency, wash durability, softness, breathability, low cost, and regulatory compliance simultaneously is a genuinely difficult multi-objective problem. Current generative models handle this better than random search, but the Pareto frontier is rarely clean. Explicit tradeoff documentation (from Paper #4) remains essential.

7.

The 80/20 for Generative AI in PFAS-Free DWR

Most coatings teams are not equipped to build GFlowNets from scratch. The practical question is where to get the highest ROI from generative AI given available resources.

If you have Paper #4's MVD and active learning loop in place:
Add a generative pre-screening layer.
Use existing open-source tools (RDKit, open-access VAE models, MolFormer) to propose novel non-fluorinated candidates and filter by predicted toxicity before entering the wet lab. Cost: computational time and a computational chemist. Payoff: a more diverse starting library.

If you are starting from scratch:
Do not skip to generative AI.
Build the MVD and data discipline from Paper #4 first. Generative models amplify what is already known; they cannot compensate for an absence of structured performance data.

If you have regulatory pressure and a short timeline:
Use generative AI for rapid analog generation around known PFAS-free DWR scaffolds (fluorine-free polyurethanes, dendrimer waxes, silicone-polysiloxane hybrids).
This is lower-risk than de novo design and faster than brute-force screening.

If you have a long-horizon product development mandate:
Invest in a true generative pipeline.
Define your property targets, build or license a property prediction model for DWR-specific outcomes, combine with a synthesis-constrained generative model, and create a patent strategy around novel structures. This is the competitive moat of the next decade.

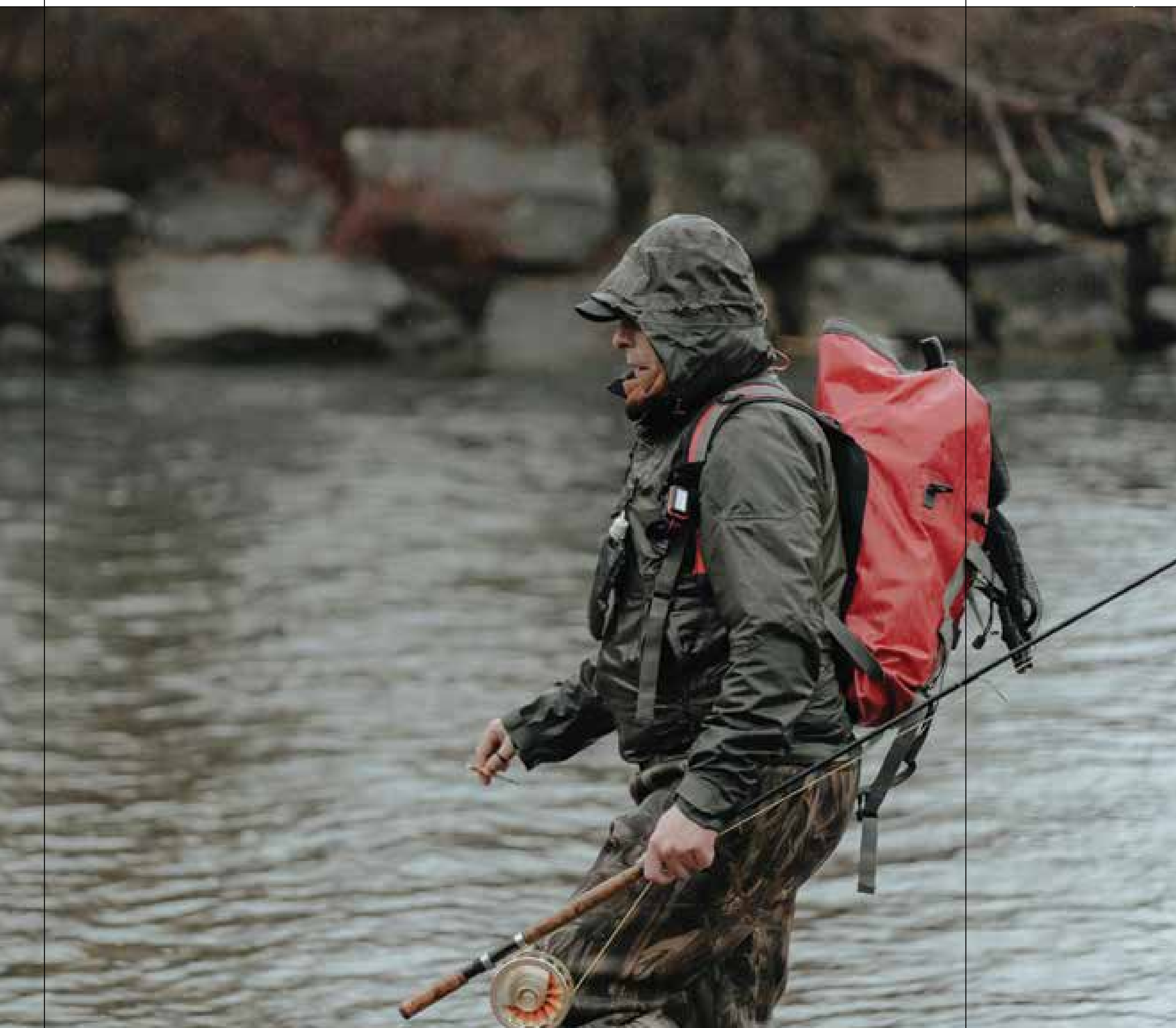
8.

Looking Forward: The 18-Month Window

The regulatory deadlines on the horizon are not soft targets.

California's 50 ppm threshold takes effect January 2027. The EU's REACH restrictions on PFAS-containing textiles and waterproofing agents are live as of January 2026. Outdoor apparel brands face their severe-wet-conditions deadline in 2028 — which, given typical R&D and mill qualification timelines, means new formulations need to be in development now.

The companies that move first on generative AI-assisted PFAS-free DWR design will do three things their competitors will not: build proprietary molecular libraries, develop process-coupled property prediction models specific to their substrate families, and accumulate structured performance data that makes every subsequent AI model more accurate.



Selected references

(1) IBM Research. MatGFN-PFAS: An AI-driven approach for toxic PFAS replacement. ACS Spring 2024.
https://research.ibm.com/publications/matgfn-pfas-an-ai-driven-approach-for-toxic-pfas-replacement

(2) Wang et al. Molecular designing of potential environmentally friendly PFAS based on deep learning and generative models. Science of the Total Environment (2024).
https://www.sciencedirect.com/science/article/abs/pii/S004896972406251X

(3) Li et al. SynFormer: Generative AI for navigating synthesizable chemical space. PNAS (2025).
https://www.pnas.org/doi/10.1073/pnas.2415665122

(4) Tom, G. et al. Self-Driving Laboratories for Chemistry and Materials Science. Chemical Reviews (2024).
https://pmc.ncbi.nlm.nih.gov/articles/PMC11363023/

(5) Szymanski, N.J. et al. An autonomous laboratory for the accelerated s ynthesis of materials. Nature (2023).
https://doi.org/10.1038/s41586-023-06734-w

(6) Alakhdar et al. Diffusion models in de novo drug design. J. Chem. Inf. Model. (2024). Review of diffusion model architectures for molecular design.

(7) MultiState. Forever Chemicals Face Sweeping Bans as States Pass PFAS Laws in 2025. January 2026.
https://www.multistate.us/insider/2026/1/22/forever-chemicals-face-sweeping-bans

(8) Trimco Group. Global PFAS Bans: Risks, Deadlines, and Textile Impact (2025).
https://www.trimco-group.com/newsroom/global-pfas-ban-regulations-and-their-impact-on-the-textile-industry

(9) Morgan Lewis. New York and California: Bans on PFAS in Textiles and Apparel Begin January 1, 2025 (2024).
https://www.morganlewis.com/pubs/2024/11/new-york-and-california-bans

(10) Arnold & Porter. Important Year-End Developments in PFAS State Law: Minnesota, New Mexico, and Connecticut (2025).
https://www.arnoldporter.com/en/perspectives/advisories/2025/12

(11) Holmquist, H. et al. Properties, performance and associated hazards of state-of-the-art durable water repellent (DWR) chemistry. Environment International (2016).
https://doi.org/10.1016/j.envint.2016.04.012

(12) Pai, S.M. et al. Machine learning applied to the design and optimization of chemical products and functional materials. (2025).
https://www.sciencedirect.com/science/article/pii/S2949822824003472

EMPEL is a PFAS-free DWR technology applied to the outer surface of textiles. Unlike traditional coatings that soak into the fabric and wash out quickly, EMPEL is pressed into the fibers – without using water or PFAS – providing long-lasting, high-performance water repellency.

Sheds water. Longer.

Contact info

4132 Jackie R d
SE Rio Rancho , NM
87124

https://empel.green/#contact