

Coach Maarten Dobbelaere	Supervisor(s) Dr. Maarten Dobbelaere Prof. Kevin Van Geem	Funding FWO
------------------------------------	--	-----------------------

AI for Decoding the Chemistry Behind Perfumes

Aim: The aim of this master thesis is to develop AI/machine-learning algorithms for molecular reconstruction of complex fragrance mixtures. By adapting petrochemical feedstock reconstruction methodologies, this thesis will use fast analytical techniques as constraints to predict detailed molecular compositions of perfumes and natural products. The results will enable the **chemical decoding** of scent profiles for artificial reconstruction.

AI-Driven Molecular Reconstruction of Fragrances

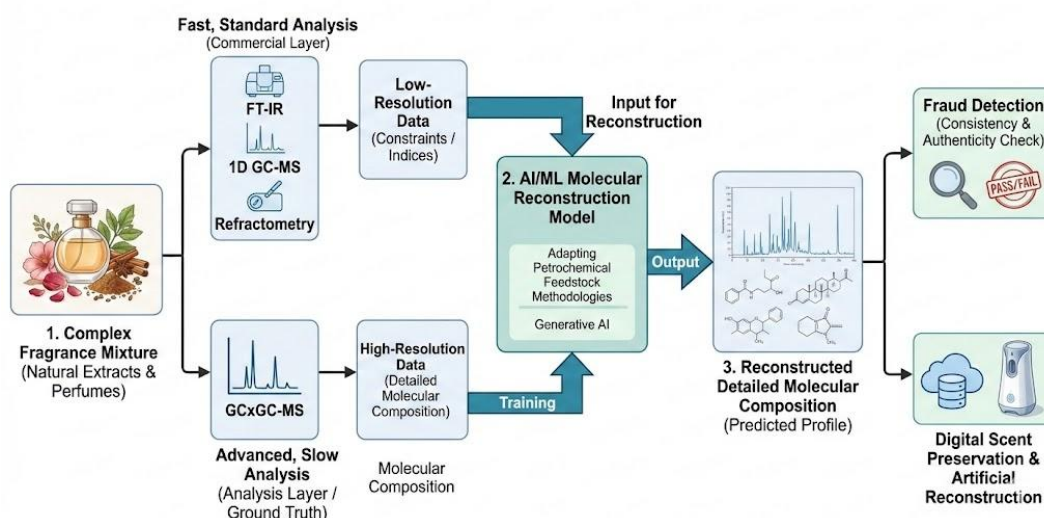


Figure 1: Proposed workflow for molecular reconstruction of the scent profile of perfumes and natural extracts.

Justification: Odorants play an important role in daily life, ranging from personal care products to insect repellents. Recently, consumer preference has shifted strongly toward natural ingredients. However, unlike synthetic chemicals, which are pure and consistent, natural extracts are chemically complex mixtures. Their composition varies significantly depending on climate, soil, and harvest conditions. Furthermore, the high monetary value of these naturals leads to sophisticated adulteration. Synthetic fillers are often added to natural products but are hard to detect using standard quality control methods.

To ensure consistency and authenticity, the detailed molecular composition of fragrance mixtures must be known. However, experimentally resolving such complex mixtures requires advanced techniques, such as two-dimensional gas chromatography (GCxGC), which are too time-consuming and expensive for routine industrial use. Consequently, the industry operates with a data gap. They possess abundant low-resolution data from fast techniques (GC-MS, FTIR, NMR), but lack the high-resolution insights needed to fully decode a scent. This challenge is identical to the feedstock reconstruction problem in petrochemistry, where complex hydrocarbon fractions must be reconstructed from limited commercial indices. Over the years, the LCT has established expertise in reconstructing complex hydrocarbon feedstocks [1-3].

The goal of this master's thesis is to adapt these petrochemical reconstruction methodologies to the field of natural products. The student will train AI models on high-fidelity GCxGC data generated at the LCT. These models will subsequently be used to reconstruct the full molecular profile of a complex fragrance using only standard, fast analytical indices as constraints. This approach enables rapid chemical decoding of scents to detect fraud and the digital preservation of scent profiles for future construction of new fragrances.

Program

- Literature survey on AI approaches for molecular reconstruction of petrochemicals and natural products
- Generation of a fragrance database
 - Digital layer: chemical data generation via quantum chemistry and machine learning
 - Commercial layer: analysis of samples using FT-IR, 1D GC-MS, and refractometry
 - Analysis layer: analysis of samples using GCxGC-MS
- Development of a generative AI algorithm to reconstruct molecular compositions from commercial data

References

1. Van Geem, K.M., et al., *Molecular reconstruction of naphtha steam cracking feedstocks based on commercial indices*. Computers & Chemical Engineering, 2007. **31**(9): p. 1020–1034.
2. Pyl, S.P., et al., *Molecular reconstruction of complex hydrocarbon mixtures: An application of principal component analysis*. AIChE Journal, 2010. **56**(12): p. 3174–3188.
3. Dobbelaere, M.R., et al., *Machine Learning for Physicochemical Property Prediction of Complex Hydrocarbon Mixtures*. Industrial & Engineering Chemistry Research, 2022. **61**(24): p. 8581–8594.