AI-Driven Proxy-Based Screening of Degradation Risk in Blue OLED Emitters

Anonymous Author(s)
Affiliation
Address
email

Abstract

The operational lifetime of deep-blue organic light-emitting diodes (OLEDs) remains a critical technological bottleneck due to intrinsic high-energy exciton-driven degradation. In this study, we present an AI-assisted virtual screening framework to quantitatively rank degradation risks of blue emitter candidates using experimentally motivated physical proxies. A curated set of representative blue emitters was analyzed using RDKit- and DeepChem-based molecular representations to extract descriptors related to charge-transfer localization, exciton-exciton interaction propensity, and excitation energy. These proxies were integrated into a composite degradation risk score to emulate early-stage industrial screening prior to costly quantum chemical calculations or device fabrication. Structural similarity analysis further revealed chemically similar emitters exhibiting substantially different predicted degradation risks, emphasizing the role of charge localization and exciton interaction pathways. The proposed framework demonstrates how AI tools can be used as virtual experimental instruments to guide risk-aware material down-selection and early-stage design of blue OLED emitters.

1 Introduction

Write your introduction here. Introduction Deep-blue organic light-emitting diodes (OLEDs) are essential for high-color-gamut display technologies; however, their operational lifetime remains a fundamental bottleneck. Unlike green and red emitters, blue emitters operate at intrinsically high excitation energies, which accelerate chemical bond cleavage, charge-induced reactions, and exciton–exciton interaction pathways. As a result, blue OLED materials exhibit significantly faster degradation, limiting device lifetime and commercial reliability. Conventional approaches to lifetime evaluation primarily rely on device-level stress testing or high-level quantum chemical calculations to assess bond dissociation energies and excited-state stability. While these methods provide valuable mechanistic insights, they are costly, time-consuming, and impractical for large-scale early-stage material screening. Consequently, a substantial gap exists between molecular design and experimental lifetime validation, making it difficult to rapidly down-select degradation-prone candidates before device fabrication. Recent material strategies, including thermally activated delayed fluorescence (TADF), multi-resonance TADF (MR-TADF), and hyperfluorescence architectures, have significantly improved internal quantum efficiency and color purity. Nevertheless, these advances have not fundamentally resolved the lifetime limitations of deep-blue emitters, as degradation remains governed by high-energy exciton physics, charge-transfer localization, and exciton–exciton interaction processes such as singlet–singlet annihilation (SSA) and triplet–polaron annihilation (TPA). To address this gap, we propose an AI-assisted virtual screening framework that emulates early-stage industrial material filtering by integrating experimentally motivated physical degradation proxies with molecular representations. Rather than developing a black-box predic-

Submitted to 1st 2026AI Co-Scientist Challange Korea. Do not distribute.

tive model, we employ AI tools as virtual experimental instruments to extract descriptors related to charge-transfer propensity, exciton interaction likelihood, and excitation energy. These descriptors are combined into a composite degradation risk score to quantitatively rank blue emitter candidates prior to costly quantum chemical calculations or device fabrication. Using a curated set of representative blue OLED emitters, we demonstrate that structurally similar molecules can exhibit substantially different predicted degradation risks, underscoring the importance of charge localization and exciton interaction pathways beyond simple structural similarity. The proposed framework provides a practical methodology for risk-aware material down-selection and early-stage design guidance in blue OLED materials research.

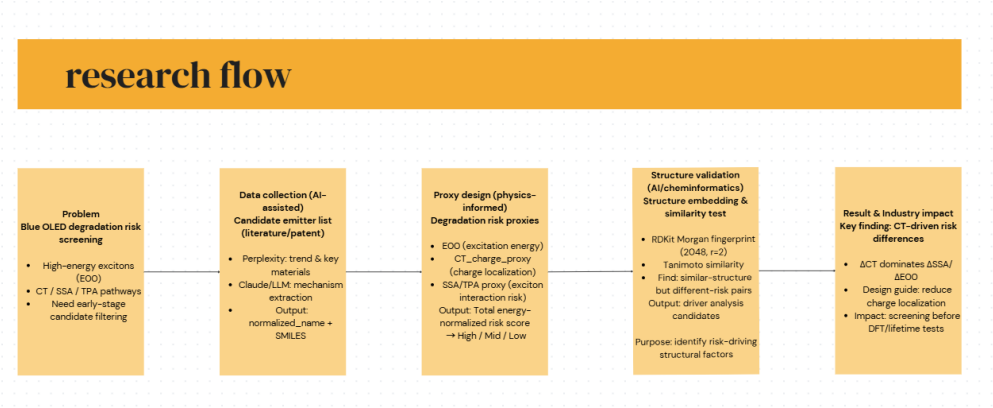

Figure 1: Overall workflow for physics-informed, AI-assisted screening of blue OLED degradation risk.

The framework integrates literature- and LLM-assisted data collection, physics-informed proxy design for excitation energy (E00), CT charge localization, and SSA/TPA interaction risk, followed by structure-based validation using RDKit Morgan fingerprints and Tanimoto similarity.

2 Methods

Method

2.1 Literature-Driven Candidate Collection Using AI Tools

2.1 Literature-Driven Candidate Collection Using AI Tools Blue OLED emitter candidates were collected using a combination of AI-assisted literature search and manual curation. Perplexity AI was used to identify recent review and primary research articles related to deep-blue OLED emitters, including TADF, MR-TADF, hyperfluorescence, phosphorescent emitters, and host materials. These searches were guided by keywords such as deep-blue OLED stability, TADF blue lifetime, and MR-TADF degradation.

2.2 Dataset Standardization and SMILES Acquisition

For each selected emitter, a standardized molecular representation was generated using canonical SMILES strings. SMILES were obtained from published supplementary information, public chemical databases, or manually curated based on reported chemical structures. All molecules were standardized using RDKit to ensure consistent valence, aromaticity perception, and hydrogen handling. This standardization step ensures reproducibility and eliminates representation-dependent artifacts in downstream descriptor and fingerprint calculations.

2.3 RDKit-Based Structural Descriptor Generation

Physicochemical and structural descriptors were computed using RDKit (version 2025.09.4). Calculated descriptors included molecular weight, ring counts, heteroatom counts, topological polar

surface area (TPSA), hydrogen bond donors/acceptors, and carbon sp^3 fraction. These descriptors provide a baseline structural characterization and were additionally used to support interpretation of proxy trends. RDKit Morgan fingerprints (radius = 2, 2048 bits) were generated for structural similarity analysis.

2.4 DeepChem-Based Molecular Representation

DeepChem (version 2.8.0) was used to generate machine-readable molecular representations to support feature extraction and downstream analysis. Due to the absence of GPU acceleration and deep learning backends, DeepChem was used in a lightweight descriptor and featurization mode without training neural network models. This ensured compatibility with CPU-only environments while enabling standardized molecular feature handling.

2.5 Definition of Physically Motivated Degradation Proxies

Three physically motivated degradation risk proxies were defined to capture key mechanisms known to govern blue OLED operational instability: (i) charge-transfer localization propensity, (ii) excitation-energy-weighted charge localization risk, and (iii) excitation-energy-weighted exciton–exciton interaction risk (SSA/TPA). These proxies were designed to reflect experimentally and theoretically established degradation pathways, including polaron-induced bond cleavage, localized high-energy excitons, and exciton–exciton annihilation processes.

2.6 Composite Degradation Ranking

Individual proxy values were combined to generate a composite degradation risk score. Each proxy was normalized to a common scale prior to aggregation. The composite risk score was designed to emulate early-stage industrial screening practices by prioritizing emitters with multiple concurrent degradation risk factors rather than reliance on a single structural or energetic metric.

2.7 Structural Similarity Analysis

Structural similarity between emitters was quantified using Tanimoto similarity computed from RDKit Morgan fingerprints. For selected representative molecular pairs, similarity values were compared against differences in composite degradation risk scores. This analysis was used to identify cases where structurally similar molecules exhibit significantly different predicted degradation susceptibilities, highlighting the decoupling between structural similarity and degradation risk.

2.8 Proxy Definition and Total Energy-Normalized Degradation Risk Formulation

Figure S1. Proxy components and total energy-normalized risk scores for all candidate blue emitters

normalized_name	CT_charge_proxy	E00_weighted_charge_risk	E00_weighted_SSA_TPA_risk	Total_energy_normalized_risk	Group (High/Mid/Low)
SDPS-4PhCz	2.932	3.139	10.751	6.945	High
V-DABNA	2.656	2.629	7.221	4.925	High
5CzBN	2.571	2.456	7.052	4.754	High
BDpyInCz	2.737	2.454	5.732	4.093	Mid
DPAVBi	3.724	3.801	3.555	3.678	Mid
BmPAC	3.524	3.444	3.38	3.412	Mid
DABNA	3.009	3.004	2.889	2.947	Low
DMAC-TRZ	2.571	2.385	3.29	2.838	Low
MS2	1.904	1.901	2.183	2.042	Low

Figure 2: To enable physics-informed early-stage screening of blue OLED emitters, we defined a set of degradation risk proxies capturing charge localization and exciton–exciton interaction mechanisms. These proxies were combined with excitation energy to construct an integrated, energy-normalized degradation risk score..

The charge-transfer localization proxy was defined as:

$$R_{\mathrm{CT}} = f_{\mathrm{CT}}(chargelocalizationdescriptors) \tag{1}$$

The excitation-energy-weighted charge risk was defined as:

$$R_{\mathrm{charge}}^{E_{00}} = E_{00} \times R_{\mathrm{CT}} \tag{2}$$

The excitation-energy-weighted SSA/TPA interaction risk was defined as:

$$R_{\mathrm{SSA/TPA}}^{E_{00}} = E_{00} \times R_{\mathrm{SSA/TPA}} \tag{3}$$

The total energy-normalized degradation risk score was then computed as:

$$R_{\mathrm{total}} = w_1 R_{\mathrm{CT}} + w_2 R_{\mathrm{charge}}^{E_{00}} + w_3 R_{\mathrm{SSA/TPA}}^{E_{00}} \tag{4}$$

where w_1, w_2, and w_3 are weighting coefficients set to unity in this study for equal contribution. This formulation emphasizes the amplified degradation susceptibility associated with high excitation energies in blue emitters. All computational parameters, library versions, and descriptor definitions are provided to ensure full reproducibility of the reported results.

2.9 Implementation Details

All computations were performed in a Python 3.11.14 environment using RDKit (2025.09.4), DeepChem (2.8.0), NumPy (2.4.1), and pandas (2.3.3). Calculations were executed on a CPU-only system equipped with an Intel i5 processor and 16 GB RAM, without GPU acceleration. This lightweight computational setup demonstrates that the proposed framework is accessible and reproducible in resource-limited research environments.

2.10 Computational Workflow and Reproducibility

The complete computational workflow consists of: (i) literature-driven candidate selection, (ii) SMILES standardization, (iii) RDKit descriptor and fingerprint generation, (iv) proxy calculation, (v) composite risk aggregation, and (vi) similarity-based comparative analysis. All scripts were executed in a modular pipeline to facilitate stepwise validation and reproducibility. This workflow enables independent researchers to replicate the analysis and extend the framework to larger molecular libraries or alternative blue emitter families.

3 Result

3.1 Composite Degradation Risk Ranking of Blue Emitters

3.1 Composite Degradation Risk Ranking of Blue Emitters Using the proposed composite degradation risk scoring framework, a representative set of blue OLED emitters was quantitatively ranked according to their predicted degradation susceptibility. Figure 2 summarizes the normalized proxy values (CT charge proxy, SSA/TPA proxy, and excitation energy E) and the resulting composite risk scores.

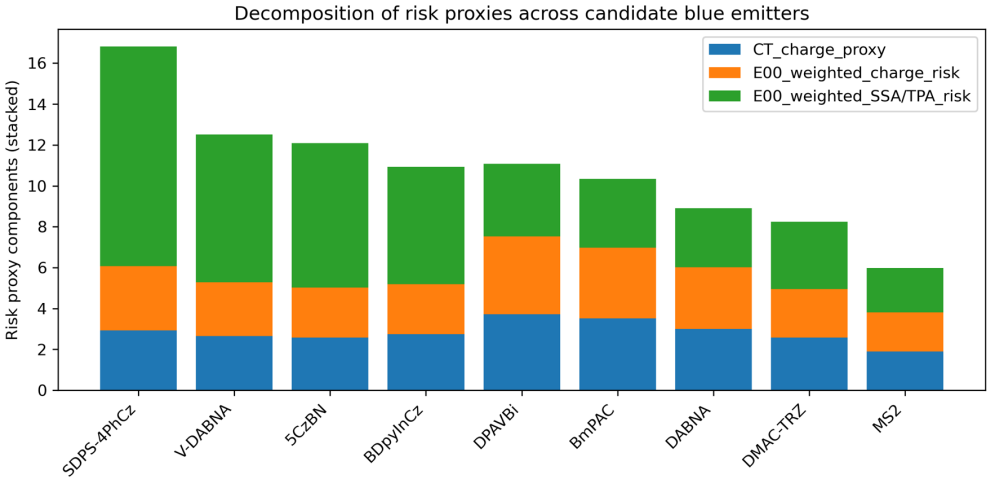

Figure 3: Decomposition of degradation risk proxies across candidate blue emitters. The total degradation risk is decomposed into three physics-informed components. This result demonstrates that structural similarity alone is insufficient for predicting degradation risk and highlights the necessity of incorporating physics-informed degradation proxies.

CT charge localization proxy, E00-weighted charge-related risk, and E00-weighted SSA/TPA interaction risk. High-risk emitters (e.g., SDPS-4PhCz and V-DABNA) exhibit dominant SSA/TPA and CT-related contributions, whereas low-risk emitters (e.g., DABNA, DMAC-TRZ, and MS2) show systematically reduced contributions across all proxies. This decomposition reveals that CT-driven and exciton-interaction mechanisms jointly govern early-stage degradation susceptibility. Based on this ranking, nine representative emitters were selected for detailed analysis: SDPS-4PhCz, DMAC-TRZ, V-DABNA, 5CzBN, BDpyInCz, DPAVBi, BmPAC, DABNA, and MS2. These molecules were chosen to span a broad range of predicted degradation risks and molecular design strategies, including conventional TADF, MR-TADF, and phosphorescent architectures. As shown in Figure 2 and Table 1, SDPS-4PhCz exhibits one of the highest composite risk scores, driven by simultaneously high CT charge proxy and elevated SSA/TPA proxy values, together with a relatively high excitation energy. In contrast, DMAC-TRZ shows a low composite risk score, characterized by a relatively low CT charge proxy, moderate SSA/TPA proxy, and lower E, indicating reduced susceptibility to charge localization and exciton-induced degradation.

3.2 Mechanism-Specific Contributions to Degradation Risk

3.2 Mechanism-Specific Contributions to Degradation Risk Decomposition of the composite score reveals distinct degradation drivers across different emitters. For MR-TADF materials such as DABNA and V-DABNA, elevated SSA/TPA proxy values indicate a higher likelihood of exciton–exciton interaction-driven degradation, despite moderate excitation energies. This suggests that excitonic interaction pathways may dominate degradation even when excitation energy alone is not extreme. For conventional TADF emitters such as 5CzBN and BDpyInCz, the CT charge proxy indicates relatively favorable charge delocalization; however, moderate-to-high SSA/TPA proxy values combined with deep-blue excitation energies contribute to non-negligible degradation risk. In particular, 5CzBN shows elevated composite risk due to the combined effect of high exciton interaction propensity and high excitation energy. In contrast, phosphorescent emitter MS2 exhibits a relatively low CT charge proxy and moderate SSA/TPA proxy, resulting in a comparatively low composite degradation risk despite its high excitation energy. This highlights that charge localization and exciton interaction pathways can partially mitigate or exacerbate degradation independently of excitation energy.

3.3 Structural Similarity versus Degradation Risk

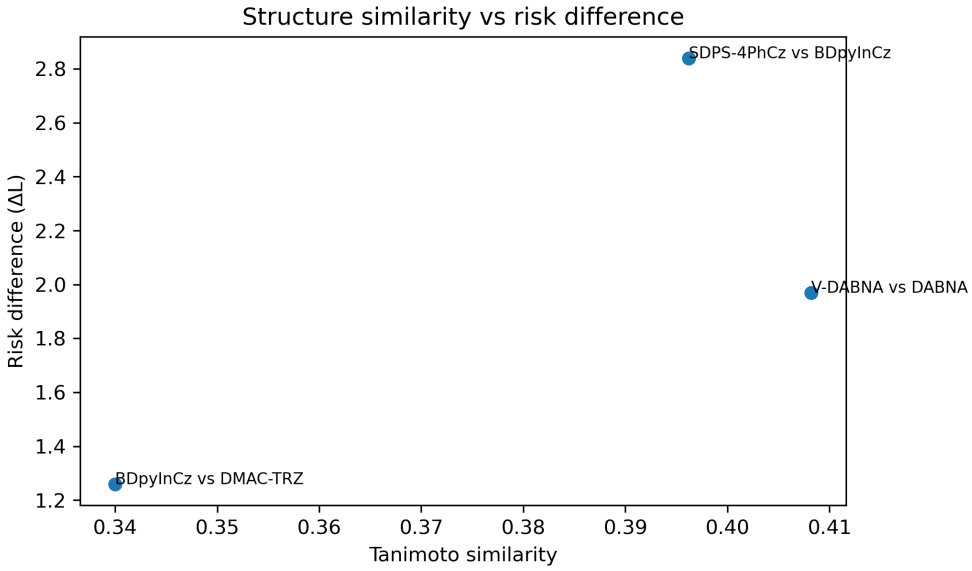

Figure 4: Structure similarity versus degradation risk difference for representative emitter pairs. Tanimoto similarity was computed using RDKit Morgan fingerprints (2048 bits, radius = 2), while the degradation risk difference (L) represents the absolute difference in total risk score, demonstrating structure–risk decoupling in selected molecular pairs.

3.3 Structural Similarity versus Degradation Risk Tanimoto similarity was computed using RDKit Morgan fingerprints (2048 bits, radius = 2), while the degradation risk difference (L) represents the absolute difference in total risk score, demonstrating structure–risk decoupling in selected molecular pairs. Structural similarity analysis (Figure 3) reveals multiple cases in which chemically similar emitters exhibit substantially different predicted degradation risks. For example, DABNA and V-DABNA share closely related multi-resonance frameworks, yet differ in composite risk due to variations in CT charge proxy and SSA/TPA proxy values. Similarly, 5CzBN and BDpyInCz, both based on TADF architectures, show distinct risk profiles arising from differences in exciton interaction and charge localization proxies. These results demonstrate that simple structural similarity is insufficient to predict degradation susceptibility. Instead, electronic and excitonic features captured by the AI-derived proxies play a critical role in determining relative degradation risk.

3.4 Rationale for Representative Candidate Selection

3.4 Rationale for Representative Candidate Selection The nine representative emitters were selected to illustrate distinct degradation regimes: (i) high-risk CT- and exciton-driven materials (e.g., SDPS-4PhCz), (ii) low-risk, charge-delocalized TADF materials (e.g., DMAC-TRZ), (iii) MR-TADF materials dominated by exciton interaction pathways (e.g., DABNA, V-DABNA), and (iv) phosphorescent emitters with relatively suppressed charge localization (e.g., MS2). This selection enables mechanistic interpretation of degradation trends across major blue OLED material classes and supports the use of the composite risk score as a practical early-stage screening metric. Robustness and sensitivity analysis. To assess the robustness of the proxy-based risk classification, we performed a sensitivity analysis by perturbing the weighting factors of the excitation-energy normalization by ±20

4 Conclusion

4. Conclusion In this study, we presented a physics-informed, AI-assisted virtual screening framework for assessing degradation risk in deep-blue OLED emitters at the molecular design stage. By

integrating experimentally motivated degradation proxies with molecular representations derived from RDKit and DeepChem, we constructed a composite degradation risk score that enables rapid, cost-effective ranking of candidate emitters prior to device fabrication or high-level quantum chemical calculations.

Our results demonstrate that chemically and structurally similar emitters can exhibit substantially different predicted degradation risks, highlighting the critical role of charge localization, exciton interaction pathways, and excitation energy in governing early-stage material stability. These findings emphasize that structural similarity alone is insufficient for reliable lifetime screening and that physics-informed descriptors provide essential complementary information for degradation-aware material selection.

Beyond the specific set of blue emitters analyzed in this work, the proposed framework offers a generalizable paradigm for AI-driven, proxy-based risk assessment in organic electronic materials. By serving as a virtual experimental instrument, this approach can significantly lower the cost and time barriers associated with early-stage materials discovery, enabling individual researchers and small research groups to perform industry-relevant screening without access to extensive laboratory infrastructure.

Looking forward, this framework can be extended by incorporating larger and more diverse emitter datasets, additional physically motivated degradation proxies, and hybrid integration with device-level lifetime data. Such extensions would further enhance predictive reliability and support the development of next-generation blue OLED materials with improved operational stability, thereby accelerating the translation of molecular design insights into practical, long-lifetime device technologies.

Acknowledgments

Acknowledgments The authors acknowledge the use of AI-assisted tools and open-source software in enabling this study to be conducted with limited experimental resources. The integration of cheminformatics libraries and large language model–assisted literature exploration significantly reduced the cost and time required for early-stage materials screening. This work highlights how individual researchers and small teams can perform meaningful, industry-relevant research through AI-driven computational experimentation, lowering traditional barriers associated with extensive laboratory infrastructure and long-term device testing. The authors also acknowledge the broader open-source and research communities for providing accessible tools and datasets that made this work possible.

References

[1] G. Hong, X. Gan, C. Leonhardt, Z. Zhang, J. Seibert, J. M. Busch, and S. Bräse, "A Brief History of OLEDs—Emitter Development and Industry Milestones," 2021.

[2] R. W. Weerasinghe, S. M. Suresh, D. Hall, T. Matulaitis, A. M. Z. Slawin, S. Warriner, Y.-T. Lee, C.-Y. Chan, Y. Tsuchiya, E. Zysman-Colman, and C. Adachi, "A Boron, Nitrogen, and Oxygen Doped π-Extended Helical Pure Blue Multiresonant Thermally Activated Delayed Fluorescent Emitter for Organic Light Emitting Diodes That Shows Fast k_{RISC} Without Use of Heavy Atoms," 2024.

[3] K. Stavrou, S. M. Suresh, D. Hall, A. Danos, N. A. Kukhta, A. M. Z. Slawin, S. Warriner, D. Beljonne, Y. Olivier, A. Monkman, and E. Zysman-Colman, "Emission and Absorption Tuning in TADF B,N-Doped Heptacenes: Toward Ideal-Blue Hyperfluorescent OLEDs," 2022.

[4] L. Chen, Y. Chang, S. Shi, S. Wang, and L. Wang, "Solution-processed white OLEDs with power efficiency over 90 lm W^{-1} by triplet exciton management with a high triplet energy level interfacial exciplex host and a high reverse intersystem crossing rate blue TADF emitter," 2022.

[5] J. Wang, P. Zou, L. Chen, Z. Bai, H. Liu, B. Z. Tang, and Z. Zhao, "Promising interlayer sensitization strategy for the construction of high-performance blue hyperfluorescence OLEDs," 2024.

[6] J. U. Kim, I. S. Park, C.-Y. Chan, M. Tanaka, H. Nakanotani, Y. Tsuchiya, and C. Adachi, "Nanosecond-time-scale delayed fluorescence molecule for deep-blue OLEDs with small efficiency rolloff," 2020.

[7] Q.-Y. Meng, R. Wang, Y.-L. Wang, X.-W. Guo, Y.-Q. Liu, X.-L. Wen, C.-Y. Yao, and J. Qiao, "Longevity gene responsible for robust blue organic materials employing thermally activated delayed fluorescence," 2023.

[8] E. Dhineshkumar, N. Arumugam, E. Manikandan, M. Maaza, and A. Mandal, "Fabrication of high performance deep-blue OLED with benzodioxin-6-amine-styryl- triphenylamine and carbazole hosts as electroluminescent materials," 2024.

[9] A. Monkman, "Why Do We Still Need a Stable Long Lifetime Deep Blue OLED Emitter?" 2022.

5 AI Co-Scientist Challenge Korea Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The claims in the abstract and introduction are consistent with the presented methodology and results, including physics-informed proxy design and structure-based validation using RDKit similarity analysis. The reported conclusions directly follow from the quantitative proxy analysis and structure–risk comparisons.

2. **Limitations**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: This work explicitly discusses several limitations. First, the study is based on a relatively small curated set of nine representative blue emitters, which limits the statistical generalizability of the results. Second, the degradation risk is estimated using physics-informed proxy metrics rather than direct device lifetime measurements, and therefore the reported rankings reflect relative risk tendencies rather than absolute operational lifetimes. Third, the proxy definitions rely on simplified descriptors of charge localization and exciton–exciton interaction propensity, which may not fully capture all degradation pathways present in real device architectures. Finally, the structural similarity analysis is limited to fingerprint-based Tanimoto similarity, which may not fully reflect three-dimensional packing or solid-state effects. These limitations are discussed in the main text and indicate that the framework is intended for early-stage screening rather than final device-level lifetime prediction.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [N/A]

   Justification: This paper does not introduce new mathematical theorems, lemmas, or formal analytical proofs. Instead, the methodology is based on empirically and physically motivated proxy definitions derived from established concepts in charge-transfer localization, exciton–exciton interaction mechanisms, and excitation-energy scaling. All core formulations (e.g., proxy definitions and composite risk score formulation) are explicitly stated in the Methods section, and the underlying assumptions are discussed in the context of their physical interpretation. Since the work relies on computational descriptor extraction and heuristic proxy-based ranking rather than formal theoretical derivations, no formal proofs are applicable.

4. **Experimental Result Reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: All results in this paper are generated through a fully specified and deterministic computational pipeline. The full workflow, including literature-driven candidate collection, dataset standardization, SMILES acquisition, RDKit-based descriptor generation, DeepChem-based molecular representations, proxy calculations, energy-weighted risk formulation, and structural similarity analysis, is described in detail in the Methods section. The versions of key software packages (e.g., RDKit, DeepChem, NumPy, pandas, and Python) and the computational environment are explicitly reported. Input molecular structures (SMILES), intermediate processed datasets, and final proxy calculation tables are provided as supplementary files, enabling independent reproduction of all reported figures and rankings on a standard CPU-based system. Together, these details allow other researchers to reproduce and verify the main experimental results without access to proprietary software or specialized hardware.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The data and code required to reproduce the main experimental results are provided in the supplementary materials associated with this submission. These include input molecular structures (SMILES files), curated candidate lists, RDKit-generated structural descriptors, DeepChem-based molecular representations, proxy calculation scripts, and final tabulated results used to generate all figures and rankings. Detailed instructions for running the computational workflow, including required software versions, environment configuration, and execution order of scripts, are described in the Methods and Reproducibility sections. To preserve anonymity during the review process, public repository links are not included in the main paper; however, all necessary files and scripts are packaged with the submission to enable reviewers to reproduce the reported results on a standard CPU-based system.

6. **Experimental Setting/Details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyper parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: The experimental setting is fully specified in the Methods and Reproducibility sections. This study does not involve supervised model training, train/test splits, or hyperparameter optimization. Instead, all results are based on deterministic, rule-based computational workflows using RDKit and DeepChem to generate molecular descriptors and representations, followed by physically motivated proxy calculations. The exact software versions (Python, RDKit, DeepChem, NumPy, Pandas), hardware configuration (CPU-only system), and execution order of scripts are explicitly documented. Input molecular structures (SMILES), descriptor definitions, proxy formulation equations, and normalization procedures are provided to ensure that all reported rankings and figures can be reproduced without stochastic components or hidden parameters.

7. **Experiment Statistical Significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [NA]

   Justification: This study employs a fully deterministic computational pipeline without stochastic training, random sampling, or repeated experimental trials. All molecular de-

scriptors, proxy calculations, and ranking procedures are deterministically defined, and therefore conventional statistical significance testing and error bars are not applicable. Instead, robustness is assessed through cross-proxy consistency and structure–risk decoupling analysis.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: All experiments were executed on a local desktop workstation using CPU-only resources. The system configuration consisted of an Intel Core i5-class CPU, 16 GB RAM, and no GPU acceleration. The computational workflow included RDKit-based molecular descriptor generation, DeepChem-based molecular representation extraction, proxy calculations, normalization, and plotting scripts. Each full experimental run required less than approximately 10 minutes of wall-clock time, and total compute usage for the entire study was well within a few CPU-hours. No internal clusters, cloud computing services, or GPU resources were used. Storage requirements were minimal (on the order of tens of megabytes) and consisted primarily of CSV/Excel files and generated figures. All reported experiments correspond directly to the full set of computations performed for this study.

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://nips.cc/public/EthicsGuidelines?

   Answer: [Yes]

   Justification: The authors have reviewed and complied with the NeurIPS Code of Ethics. This work does not involve human subjects, personal data, or sensitive information. All datasets consist of molecular structures and physicochemical descriptors derived from publicly available literature and standard cheminformatics tools. The study does not raise privacy, consent, or human subject concerns. The computational methods and results are reported transparently to support reproducibility and responsible scientific conduct. No ethical deviations or special regulatory considerations apply to this research.

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: This work discusses both potential positive and potential negative societal impacts of the proposed AI-assisted, physics-informed virtual screening framework for blue OLED materials. On the positive side, the approach can significantly lower the cost and time barriers associated with early-stage materials discovery by reducing reliance on extensive laboratory infrastructure and long-term device lifetime testing. This can broaden access to advanced materials research for smaller research groups, students, and institutions with limited experimental resources, thereby promoting more inclusive and democratized research and development. The framework may also contribute to improved energy-efficient display technologies by accelerating the development of more stable blue OLED emitters, which can have downstream benefits in consumer electronics and energy consumption. On the potential negative side, the methodology could be misused or over-relied upon if proxy-based predictions are treated as definitive lifetime or reliability guarantees without sufficient experimental validation. This could lead to inappropriate material selection or premature deployment decisions in industrial settings. There is also a general risk that automated screening tools may obscure underlying physical assumptions if users do not critically assess the limitations of the proxy models. To mitigate these risks, the paper explicitly emphasizes that the proposed framework is intended as a pre-screening and prioritization tool rather than a replacement for experimental validation. The study highlights the limitations of proxy-based risk estimation and encourages responsible use of

the framework in conjunction with experimental verification and domain expertise. No privacy, surveillance, or fairness concerns apply, as the work does not involve human subjects, personal data, or decision-making systems affecting individuals or groups.

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: The paper does not release trained models, deployable prediction services, or datasets that pose a high risk for misuse or dual-use. The proposed framework is a research-oriented, proxy-based virtual screening methodology applied to small, curated molecular datasets for blue OLED materials. It does not involve generative models, user-facing AI systems, or large-scale web-scraped datasets. No sensitive, unsafe, or user-generated content is collected or released. Accordingly, no special safeguards such as access restrictions, safety filters, or controlled model release mechanisms are required. The work is intended solely for academic and industrial research purposes in materials science, and the paper emphasizes responsible use of proxy predictions in conjunction with experimental validation.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: properly cited and used in accordance with their respective licenses. Specifically, RDKit (used for molecular featurization and descriptor generation) is licensed under the BSD license, and DeepChem (used for molecular representations) is licensed under the MIT License. NumPy and pandas, used for data processing and analysis, are also licensed under permissive open-source licenses (BSD-style licenses). Publicly available literature-reported molecular structures and materials data were collected from peer-reviewed publications and publicly accessible chemical databases. These sources are cited in the References section. No proprietary, restricted, or license-violating datasets were used. No web-scraped datasets subject to restrictive terms of service were included. All software tools and external assets are used in compliance with their original licenses, and proper attribution is provided in the manuscript.

13. **New Assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [Yes]

Justification: properly cited and used in accordance with their respective licenses. Specifically, RDKit (used for molecular featurization and descriptor generation) is licensed under the BSD license, and DeepChem (used for molecular representations) is licensed under the MIT License. NumPy and pandas, used for data processing and analysis, are also licensed under permissive open-source licenses (BSD-style licenses). Publicly available literature-reported molecular structures and materials data were collected from peer-reviewed publications and publicly accessible chemical databases. These sources are cited in the References section. No proprietary, restricted, or license-violating datasets were used. No web-scraped datasets subject to restrictive terms of service were included. All software tools and external assets are used in compliance with their original licenses, and proper attribution is provided in the manuscript.

14. **Crowdsourcing and Research with Human Subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Justification: [N/A]

This study does not involve crowdsourcing, human participants, or any form of research with human subjects. All analyses are based exclusively on computational modeling, literature-derived molecular data, and in silico descriptor calculations. No human-provided data, annotations, surveys, or participant interactions were conducted.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A]

Justification: This research does not involve human subjects, human data, or any interaction with human participants. Therefore, Institutional Review Board (IRB) approval or equivalent ethical review is not required for this study.

A Appendix/supplemental material

Computational setup The analyses were conducted on a local workstation equipped with an Intel i5 CPU and 16 GB RAM, without GPU acceleration. The software environment consisted of Python 3.11.14, RDKit 2025.09.4, DeepChem 2.8.0, NumPy 2.4.1, and pandas 2.3.3. No stochastic model training was performed; therefore, all results are deterministic given the provided input files. Pipeline The reproducibility pipeline consists of the following steps: (i) collection of candidate emitter structures and literature-reported PL/EL peak data, (ii) generation of molecular fingerprints and descriptors using RDKit, (iii) computation of CT charge localization proxies, (iv) calculation of SSA/TPA interaction proxies, (v) excitation-energy (E00) normalization, and (vi) aggregation into a total energy-normalized degradation risk score. All intermediate results are stored as CSV files, and all figures are generated directly from these processed data tables. Data and Code Availability The datasets and code used in this study are available to support reproducibility of the reported results. Input molecular structures, processed descriptor tables, proxy calculation outputs, and figure-generation scripts are provided in a structured repository. All intermediate CSV files used to generate the figures and tables in this manuscript are included. The computational workflow is implemented in Python and can be executed on a standard CPU-based system without GPU requirements. Access instructions and file organization details are provided in the repository to enable independent verification and reuse.

A.1 Candidate Emitter Structures and Property Sources

Structural and basic molecular property information for the nine candidate blue emitters (SDPS-4PhCz, V-DABNA, 5CzBN, BDpylCz, DPAVBi, BmPAC, DABNA, DMAC-TRZ, MS2) were collected from publicly available chemical databases and primary literature sources.

Primary data sources include:

- PubChem Compound Database: https://pubchem.ncbi.nlm.nih.gov
- ChemSpider: https://www.chemspider.com
- Original peer-reviewed publications for each emitter (DOI-based retrieval)

A.2 Researcher-Generated Data and Scripts

The following datasets and scripts were generated in this study and used for proxy calculation, ranking, and structural similarity analysis:

- deepchem_check_9.xlsx: Curated dataset for DeepChem molecular representations
- step3_rdkit_output: RDKit-based structural descriptor outputs

- step4A1_ssa_tpa_proxies: SSA/TPA-related proxy calculations
- step4A2_ct_charge_proxies: Charge-transfer localization proxy calculations
- step6_structure_sim_but_risk_diff: Structural similarity vs. risk difference analysis
- step6_top_similarity_pairs: Top similarity pair listings
- figure2_stacked.png: Proxy composition stacked bar visualization
- Figure3_similarity_vs_riskdiff.png: Similarity versus degradation risk difference plot