Journal of Materials Chemistry A

ROYAL SOCIETY OF CHEMISTRY

PAPER

Published on 14 July 2026

Cite this: DOI: 10.1039/d6ta03136f

Validation of large language model artificial intelligence-integrated optical microscopy analysis: application to *in situ*/*operando* electrochemistry

Dogyeong Kim,†[ab] Da Hyeon Seok,†[ab] Seung-Ho Yu,[b] Jun Hyuk Moon,[b] Jialu Wang,[a] Jai Hyun Koh,[ae] Chansol Kim,[ae] Jae-Young Choi,[cd] Hyung-Suk Oh *[acd] and Woong Hee Lee *[ae]

Analyzing chemical states and monitoring their transformations in *in situ* and *operando* environments are indispensable in chemistry and materials science. However, achieving high-resolution analysis under *operando* conditions entails substantial instrumentation costs and limited availability. Herein, we propose and validate a large language model (LLM)-based artificial intelligence (AI)-integrated optical microscopy (OM) analysis, enabling researchers to conduct analyses with ease and at minimal cost. Using LLM AI, we address the limitations of optical microscopy (OM) in performing quantitative compositional analysis while retaining its low cost. The performance of the LLM AI-integrated OM analysis is validated by comparing results with those obtained from surface-sensitive and bulk-sensitive spectroscopy under pristine/*ex situ* conditions and *in situ*/*operando* environments. The effective analysis depth of the LLM AI-integrated OM is estimated to lie between those of the surface- and bulk-sensitive techniques, comparable to typical visible-light penetration depths. More advanced LLM AI versions are expected to further enhance analytical consistency and reproducibility. Our work can aid researchers worldwide by offering a cost-effective and accessible analytical methodology that broadens participation in the materials and chemical sciences.

Received 14th April 2026
Accepted 4th July 2026

DOI: 10.1039/d6ta03136f

rsc.li/materials-a

Introduction

In chemistry and materials science, analyzing the chemical states of materials and tracking their transformations during reactions are essential for elucidating reaction mechanisms.[1,2] Various spectroscopic techniques, such as X-ray photoelectron spectroscopy (XPS), X-ray absorption spectroscopy (XAS), near edge X-ray absorption fine structure (NEXAFS), Raman, and IR, can probe material states not only under pristine or *ex situ* conditions but also under *in situ* and *operando* environments.[3–5] However, high-resolution analyses under *operando* conditions typically require extremely expensive instrumentation or offer limited accessibility, making such studies difficult for many researchers. Optical microscopy (OM), one of the most traditional analytical methods, offers a practical solution in such cases.[6,7] The cost of OM has fallen substantially, and even models of below 50 USD can offer video recording capabilities for *in situ*/*operando* experiments. Moreover, OM provides geometric information. However, in contrast to other analytical techniques that enable quantitative compositional analysis, OM has certain limitations.

Artificial intelligence (AI) has recently received significant attention and has been increasingly integrated into numerous areas of research.[8–10] It enables faster, more efficient, and more accurate data processing and interpretation, including efficient screening to identify promising compounds and accurate prediction of mechanical, optical, thermal, and other material properties.[11–14] However, studies employing advanced AI typically require significant computational resources and collaboration with experts in the field. Large language model (LLM)-based AI models, represented by ChatGPT, demonstrate impressive performance, not only in handling language in a human-like manner but also in processing diverse data types through multimodal capabilities.[15–18] LLM-based AI is among

[a]*Clean Energy Research Center, Korea Institute of Science and Technology (KIST), Hwarang-ro 14-gil 5, Seongbuk-gu, Seoul 02792, Republic of Korea. E-mail: abcabac@kist.re.kr; hyung-suk.oh@kist.re.kr; Tel: +82 (0)2 958 5292*

[b]*Department of Chemical and Biological Engineering, Korea University, 145, Anam-ro, Seongbuk-gu, Seoul 02841, Republic of Korea*

[c]*School of Advanced Materials Science and Engineering, Sungkyunkwan University, 2066, Seobu-ro, Jangan-gu, Suwon-si, Gyeonggi-do, 16419, Republic of Korea*

[d]*KIST-SKKU Carbon-Neutral Research Center, Sungkyunkwan University, 2066 Seobu-ro, Jangan-gu, Suwon 16419, Republic of Korea*

[e]*Division of Energy and Environmental Technology, KIST School, Korea University of Science and Technology (UST), Seoul 02792, Republic of Korea*

† These authors contributed equally to this work.

the most cost-effective and readily accessible AI technologies for non-expert researchers.[19,20] Recent studies have leveraged multimodal large language models (LLMs) to automatically extract material property data by classifying and interpreting figures, spectroscopic graphs, and microscopic images in scientific literature.[21,22] However, these applications remain limited to figure recognition and morphological analysis in literature.

In this study, we aim to propose and validate a tool that combines OM with LLM- and AI-based interpretation for quantitative compositional analysis, enabling researchers to perform analyses easily and at a low cost. Although many studies have reported the use of LLM AI in research, few have applied it to analytical applications and systematically validated its performance under both pristine and *ex situ* conditions, as well as *in situ* and *operando* environments. The workflow of this study is summarized in Fig. 1. Copper was selected as the analyte because of its color variation among different chemical states. Before analyzing the unknown samples, five OM reference images for each reference state were input into ChatGPT to characterize the references without providing any other specific instructions (Fig. 1b and S1–S5). Recent LLM-based AIs are capable of performing highly sophisticated analyses by integrating comprehensive information (such as color, morphology, and texture) that often exceeds human perceptual limits. Therefore, we anticipated that the AI could successfully perform quantitative analysis by autonomously synthesizing the diverse visual information present in the photographs. Quantitative OM image analysis of three unknown samples was carried out using ChatGPT and cross-validated using surface-sensitive techniques (NEXAFS and XPS) and a bulk-sensitive technique (XAFS) (Fig. 1c and d). Extending this approach, we applied the same workflow to electrochemical *operando* experiments (Fig. 1e). Our results suggest that LLM AI-integrated OM has the potential to become a simple and cost-effective analytical methodology.

Experimental

Fabrication of electrode

The powders used as references were Cu nanoparticles, Cu_2O, CuO, $Cu(OH)_2$, and $Cu_2CO_3(OH)_2$. The electrodes for *ex situ* analysis were prepared by dispersing commercial CuO and Cu_2O powders with 5 wt% Nafion in isopropyl alcohol (IPA). The mixture was then sprayed onto a gas diffusion layer (GDL; 39BB, Sigracet) at 1 mg cm^{-2}. The electrodes for the *operando* analysis were prepared by spraying CuH (Cu hydroxide) ink onto a GDL. CuH ink was prepared by stirring $CuCl_2$ (10 mL, 0.1 M) in KOH (90 mL, 1 M) for 30 min, followed by vacuum filtration. The filtered powder (60 mg) was dispersed in a 5 wt% Nafion solution (63 mg) in IPA and sprayed using the same method.

Electrochemical test

Electrochemical experiments for *ex situ* analysis were performed using a three-electrode system with a graphite counter electrode and an Ag/AgCl reference electrode. A 1 M $KHCO_3$ solution was used as the electrolyte, and the reaction was

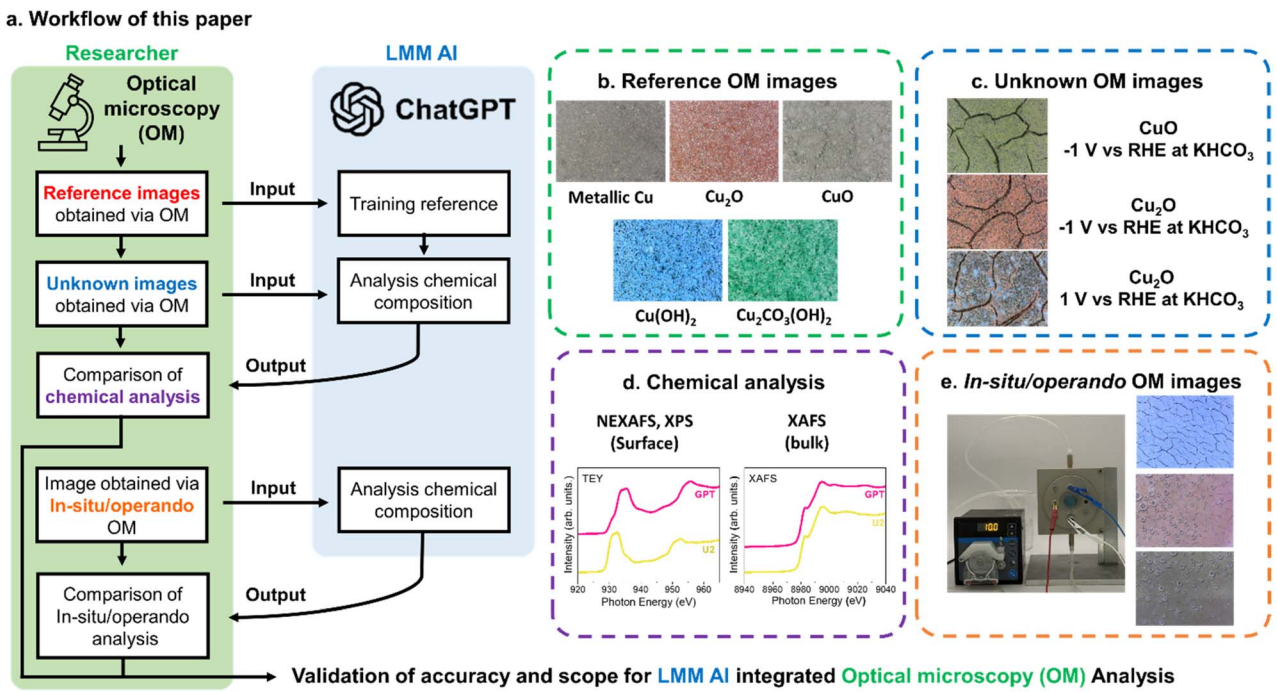

Fig. 1 Schematics of the LLM AI-integrated optical microscopy (OM) analysis: (a) workflow of this study; OM images of (b) reference samples and (c) pristine unknown samples are provided to an LMM AI, such as ChatGPT; (d) comparison of chemical compositions obtained by ChatGPT with those from experimental analyses; and (e) *in situ*/*operando* OM images of an electrode. The LMM AI-integrated OM analysis can also be applied to *in situ*/*operando* electrochemistry.

carried out for 1 h under an applied potential using a versatile potentiostat (VSP, Biologic). To obtain images for comparison with the *operando* analysis, the CuH electrode was tested under the same reaction conditions at open-circuit voltage (OCV), −0.4 V, −0.8 V, and after OCV, using the configuration in Fig. 1e.

Data collection of electrode images

The reference electrode and electrodes before and after the electrochemical reaction were examined using a USB digital microscope (1600× magnification; GL Optics). A single cell equipped with a transparent window was used to observe the surface of the CuH electrode under real-time reaction conditions. Microscopic images were captured from five randomly selected locations under identical lighting conditions. To enhance the uniformity of the visual information and minimize variations such as focus shifts or lighting changes, we carefully adjusted the focus at each step. Data consistency was assessed by calculating the mean and standard deviation of the image compositions.

Prompting GPT

Image analysis was performed using GPT-4o. Initially, five microscopic images of Cu, CuO, Cu_2O, $Cu(OH)_2$, and $Cu_2CO_3(OH)_2$ powders were provided as reference data for GPT-4o to analyze the oxidation state and composition of the electrode. Subsequently, microscopic images of the electrodes acquired after or during the electrochemical reaction were uploaded, and GPT-4o was instructed to measure the composition to one decimal place based on the five references. For model comparison, the GPT-free version and GPT-5 were also applied. In the GPT-free version, no reference images were input; electrode images were analyzed based solely on visual information such as color and texture. In GPT-5, the five reference images were pre-entered, and GPT-5 was instructed to analyze the electrode compositions based on these references. The measured results were then presented in a graph as in the workflow diagram in Fig. 1, using each reference value and the relative component ratio.

The Pearson correlation coefficient between the AI-based generated spectrum and the experimental spectrum was calculated using the following equation:

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}}$$

Physical characterization

XAS and NEXAFS were conducted at the 1D and 10D beamlines of the Pohang Accelerator Laboratory (PAL) to analyze the chemical state of the electrode. XPS was used to investigate the oxidation states of the electrode surfaces.

In situ/*operando* analyses

XAS was performed using a custom-designed electrochemical cell at the 1D and 10D beamlines of PAL. Electrochemical experiments were conducted to investigate the oxidation state of the electrode under OCV, −0.4 V, and −0.8 V. The device was stabilized before measurements at each voltage. The acquired XAFS and NEXAFS data were analyzed using the Athena software. XAS linear combination fitting was performed using Cu^0, Cu^{1+}, and Cu^{2+} as references.

Results and discussion

Validation of LLM AI-integrated OM analysis for pristine samples

Three unknown samples (U1–3) were produced by the electrochemical treatment of Cu material deposited on carbon paper, and five region-resolved OM images were obtained for each sample (Fig. 2a–c and S6–S8). The obtained OM images were input into ChatGPT for quantitative compositional analysis using the five reference materials (SI Note 2). The chemical compositions obtained by ChatGPT are shown in Fig. 2d–f. Based on the apparent color of the unknown samples, the chemical compositions derived from ChatGPT appear to be reasonably consistent with visual expectations. To further validate the ChatGPT analysis, surface-sensitive techniques (NEXAFS and XPS) and a bulk-sensitive technique (XAFS) were applied to the unknown samples. Although the spectroscopy results can be fitted using linear calibration, the presence of several spectroscopically similar references (Cu^{2+} references: CuO, $Cu(OH)_2$ and $Cu_2CO_3(OH_2)$) leads to multiple plausible solutions (Fig. S9 and S10).[23–25] Therefore, to validate the accuracy of ChatGPT, we generated spectra from the ChatGPT-derived chemical compositions and compared them with the experimental data. The surface-sensitive spectroscopies used in this study—NEXAFS (TEY), NEXAFS (TFY), and XPS—have analysis depths on the order of a few to several tens of nanometers, depending on the excitation or kinetic energy.[26,27] By comparison, bulk-sensitive XAFS reflects the overall chemical composition of the catalyst.[28–30] Pearson correlation coefficients were calculated to compare AI-generated spectra with experimental spectra more quantitatively. Interestingly, the spectra generated *via* ChatGPT are generally more similar to the surface-sensitive spectra, yet they still show reasonable agreement with both surface- and bulk-sensitive measurements (Fig. 2g–l and S11, S12). In the case of U1, excellent agreement is observed with surface-sensitive analyses, whereas U2 aligns more closely with bulk-sensitive measurements, indicating sample-to-sample variation. The AI-generated spectra showed a high degree of similarity with both the surface and bulk-sensitive analysis. (Fig. S13). These results imply that the effective analysis depth of LLM AI-integrated OM lies between those of surface- and bulk-sensitive spectroscopies.

Discussion of LLM AI-integrated OM analysis

The primary points of interest regarding LLM AI-integrated OM are the analytical performance of ChatGPT and the effective

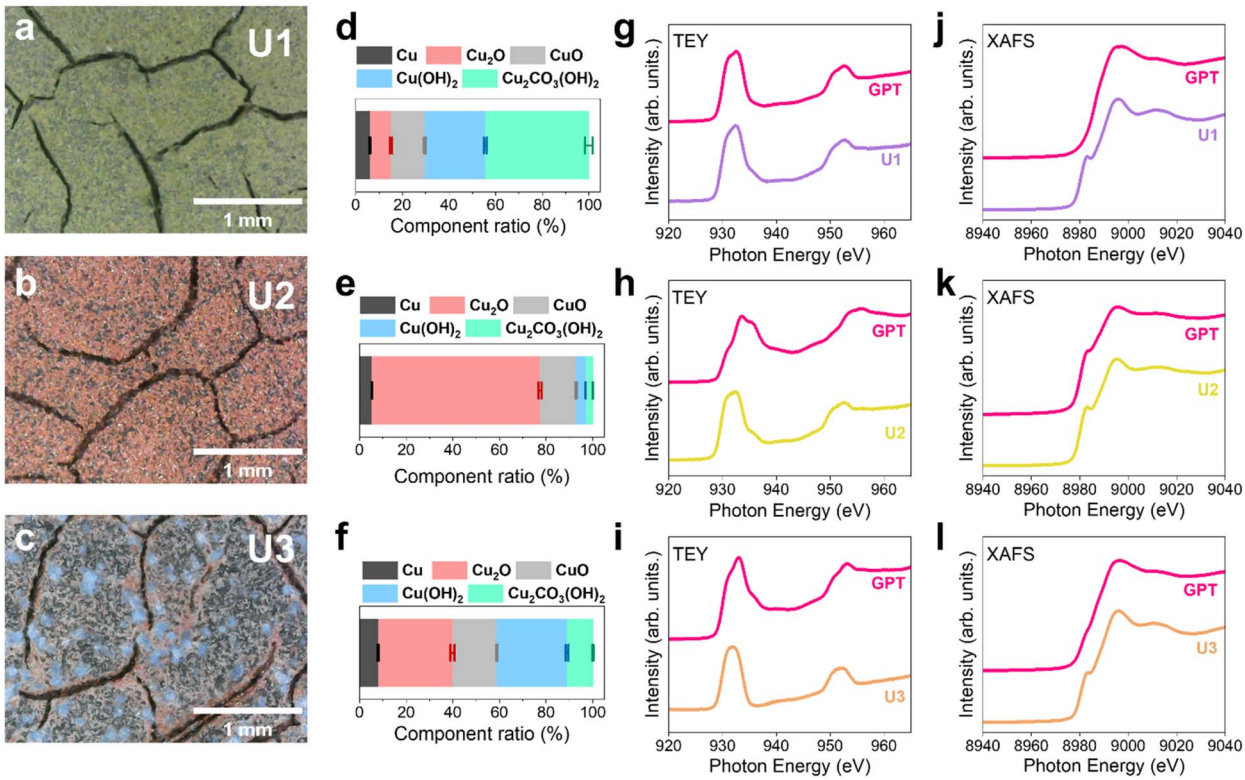

Fig. 2 Analysis of ChatGPT-integrated OM for three pristine samples: OM images of (a) U1, (b) U2, and (c) U3; (d–f) chemical compositions analyzed by ChatGPT; (g–i) NEXAFS spectra; and (j–l) XAFS spectra.

analysis depth of OM. In this section, we discuss the analysis results obtained using different ChatGPT models and examine the effective analysis depth of the OM-based method. For the three unknown samples, we compared analysis outputs from ChatGPT-4 (free, without reference inputs), ChatGPT-4o, and ChatGPT-5; the results are summarized in Fig. 3a. Small variations among the ChatGPT versions were observed, attributed to the clear color-chemistry mapping of Cu, enabling robust identification of its chemical states (Fig. S14–S17).[19] However, the standard deviation of the results—an indicator of uncertainty—varied substantially among the ChatGPT versions (Fig. 3b). The free GPT-4 version without reference inputs exhibited significantly higher standard deviations than the other versions, suggesting that providing reference images to ChatGPT is necessary to enhance analytical consistency and reproducibility. To further validate the robustness of the proposed approach, the same analysis was performed using another LLM, Gemini-3.1 Pro (Fig. S18). The results demonstrated that both surface-sensitive and bulk-sensitive spectroscopic data enabled comparable predictions of chemical states, consistent with those obtained from the ChatGPT-based analysis (Fig. S19–S22). Notably, ChatGPT-5 showed a markedly reduced dispersion relative to ChatGPT-4o, implying that more advanced model versions may provide greater analytical consistency. Interestingly, however, when examining the similarities across the different LLM versions, no significant difference was observed. (Fig. S13) the effective analysis depth of OM is governed by the visible-light penetration depth of the material. Although visible light has a much longer wavelength than hard X-rays (used for bulk analysis) and soft X-rays (used for surface analysis), its lower absorption coefficient can result in a large penetration depth, depending on the material.[31] To estimate the analysis depth in this study, we first performed linear calibration fitting using three analytical techniques with distinct information depths. We assumed effective analysis depths of 10 nm for TEY (shortest), 100 nm for TFY, and 1000 nm for bulk-sensitive XAFS (corresponding to the sample thickness). We then input the GPT-derived results to examine whether the GPT-based analysis yielded a solution between these depth models. However, because the system admits no mathematical solutions under these assumptions, the analysis depth cannot be estimated using this approach. Assuming compositional continuity with respect to position, we introduced the GPT position as an optimization variable, fitted a spline to the data, and identified the curvature-minimizing point. This value is considered an approximate estimate of the analysis depth. The estimated depths for the three unknown samples are shown in Fig. 3c–e and S23, S24. In addition, the estimated depths for the three GPT versions are summarized in Fig. 3f. The effective analysis depth of the LLM AI-integrated OM was found to be approximately 150–600 nm, comparable to typical visible-light penetration depths. Notably, the estimated penetration depth appears to be sample dependent, with U1 exhibiting an analysis depth of approximately 100–200 nm, whereas U2 displays values greater than 400 nm. This observation is attributable to differences in visible-wavelength

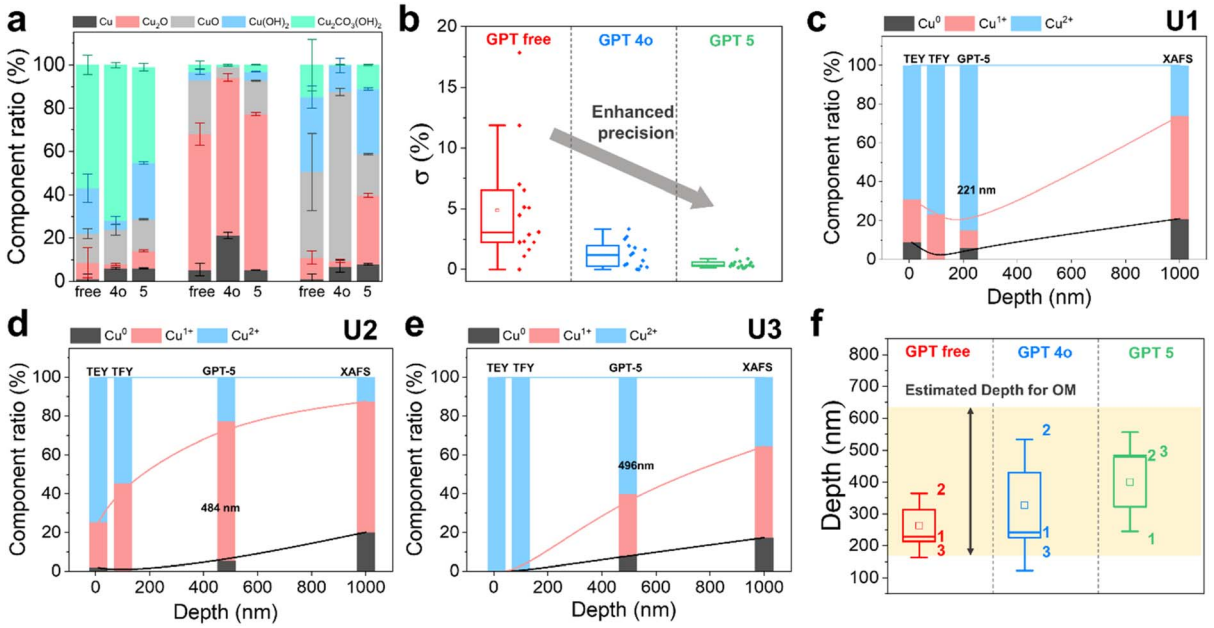

Fig. 3 Analysis of the precision and estimated analysis depth of ChatGPT: (a) chemical composition results for U1, U2, and U3 analyzed using three different ChatGPT models; (b) standard deviations for each model; analysis depth of GPT-5 estimated by spline fitting for (c) U1, (d) U2, and (e) U3; and (f) estimated analysis depths for the three ChatGPT models.

penetration depths among materials.[31] In particular, Cu_2O (unlike CuO) has a penetration depth on the order of ~1 μm, allowing optical microscopy to access deeper layers. Accordingly, the expected penetration depth scales positively with the amount of Cu_2O present (Fig. S25).[32,33] To experimentally confirm different depth by materials, the LMM analysis was applied to the fabricated two-layer electrodes (Fig. S26–28 and Table S1). For metallic Cu top layer, the signal from the bottom $Cu(OH)_2$ layer (blue bar) dropped precipitously at a loading of 0.5 mg cm^{-2}, whereas for the Cu_2O top layer, the signal from the bottom $Cu(OH)_2$ layer remains measurable even at 1.0 mg cm^{-2}. These findings establish that the visible light penetration depth (the analytical depth of our method) is inherently material-dependent, spanning a range from a few hundred nanometers to a few micrometers.

Validation of LLM AI-integrated OM analysis for *in situ*/*operando* experiments

In situ/*operando* analysis is indispensable for understanding reaction phenomena and enables direct observation of material states under reaction conditions. For *in situ*/*operando* LLM AI-integrated OM analysis, we fabricated an electrochemical flow cell that enabled optical observation of the electrode during electrochemical reactions (Fig. 4a). To conduct *in situ*/*operando* OM analysis, a cathodic potential was applied to a CuH electrode in a 1 M $KHCO_3$ electrolyte while optically monitoring the electrode (Fig. 4b and S29). Upon application of the cathodic potential, the color of the electrode changed markedly (Fig. 4c and S30–S33). Bubbling was observed following the color change, indicating that Cu reduction occurred first, followed by the hydrogen evolution reaction.[34] The *in situ*/*operando* OM images were input into GPT-5 to obtain chemical compositions, presented in Fig. 4d and S34. As the duration under cathodic potential increased, the proportion of Cu^{2+} decreased, while those of Cu^+ and metallic Cu increased, in accordance with theoretical expectations.[35] Interestingly, after OCV condition, the sample exhibits a shift toward a more metallic state. This occurs because, under a reductive current, the surface undergoes oxidation despite the applied reductive potential due to the localized alkaline pH. However, under OCV conditions, the localized pH responsible for the oxidation dissipates, allowing the surface to equilibrate with the internal reduced state. To further validate the *in situ*/*operando* LLM AI-integrated OM analysis, *in situ*/*operando* XAFS (bulk sensitive) and NEXAFS (surface sensitive) measurements were performed under similar conditions (Fig. S35). Before applying the cathodic potential, both XAFS and NEXAFS showed Cu hydroxide features, consistent with the LLM AI-integrated OM analysis. Upon application of the cathodic potential, the *in situ*/*operando* XAFS spectrum transformed to a metallic profile, while the *in situ*/*operando* NEXAFS spectrum exhibited a mixture of Cu^{2+}, Cu^+ and metallic Cu. Spectra generated from the GPT-derived compositions were compared with the experimental spectra for detailed comparison (Fig. 4e and f). Relative to XAFS, the generated spectra exhibit less pronounced metallic features, whereas relative to NEXAFS, the Cu^{2+}-related peak is less prominent. These differences suggest that LLM AI-integrated OM analysis possesses an intermediate effective analysis depth between surface- and bulk-sensitive techniques. Collectively, these findings validate LLM AI-integrated OM quantification for both pristine catalysts and electrochemical *in situ*/*operando* measurements.

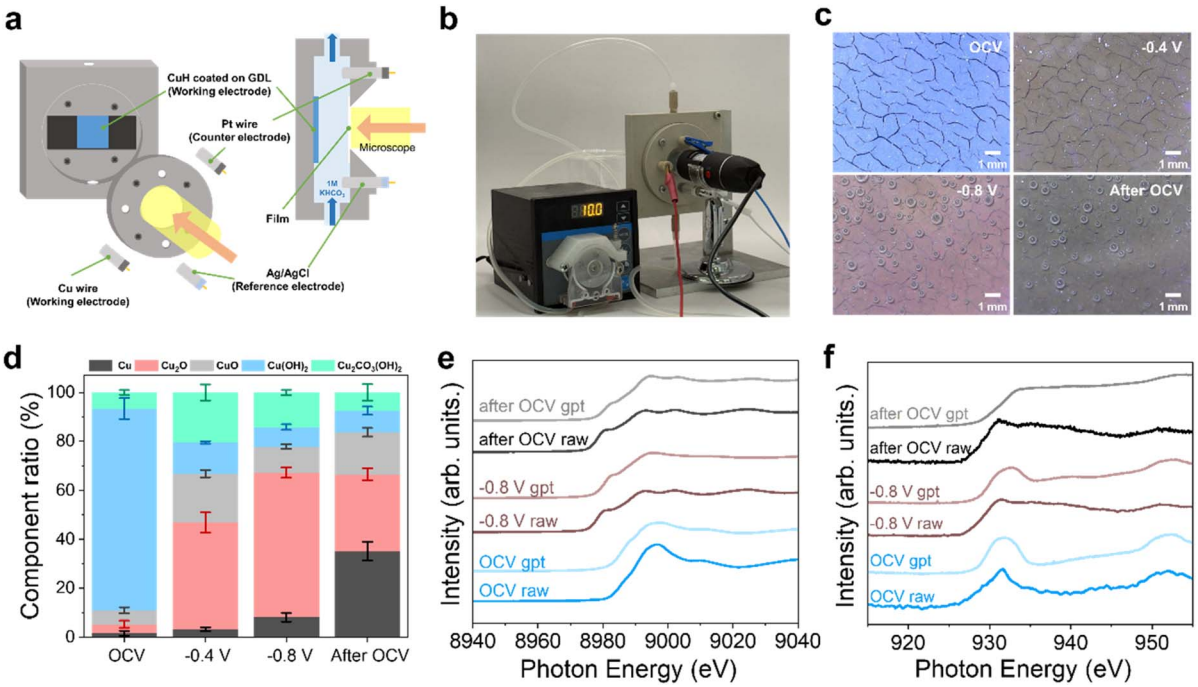

Fig. 4 *In situ*/*operando* analysis of a CuH electrode: (a) schematic illustration and (b) photograph of the electrochemical cell; (c) OM images of the CuH electrode at different cathodic potentials; (d) chemical compositions analyzed by ChatGPT; (e) *in situ*/*operando* Cu K-edge XAFS spectra; and (f) *in situ*/*operando* Cu L-edge NEXAFS spectra.

Conclusion

In summary, we validated the quantification results of the LLM AI-integrated OM analysis by comparing them with results from surface- and bulk-sensitive spectroscopies. For the three unknown samples, LLM AI-integrated OM aligned closely with surface-sensitive spectroscopy while exhibiting partial agreement with bulk-sensitive spectroscopy. These results imply that LLM AI-integrated OM analysis exhibits good performance in characterizing unknown materials. Providing reference data prior to GPT analysis substantially reduced the standard deviation of the quantification results, demonstrating that the input of reference images is necessary to enhance accuracy and reproducibility. Furthermore, the advanced GPT-5 model exhibits a lower standard deviation than the GPT-4 model, suggesting that advancements in LLMs may enhance analytical consistency. Based on comparisons with bulk- and surface-sensitive spectroscopies, the effective analysis depth of LLM AI-integrated OM lies between those of bulk- and surface-sensitive techniques and is strongly influenced by the visible-light penetration depth of the material. Moreover, in *in situ*/*operando* electrochemical systems, LLM AI-integrated OM exhibited trends consistent with established spectroscopic measurements, thereby validating its potential as a powerful analytical tool. Building upon this validation, this validated methodology can be broadly applied across various electrochemical reactions, such as HER, OER, and CO_2RR to elucidate the correlation between real-time electrode surface state and catalytic performances. Overall, this study demonstrates that LLM AI-integrated OM can serve as a low-cost and accessible analytical methodology that contributes to research in chemistry and materials science.

Author contributions

Dogyeong Kim and Da Hyeon Seok: investigation, formal analysis, data curation, writing. Seung-Ho Yu: resources. Jun Hyuk Moon: resources. Jialu Wang: validation. Jai Hyun Koh: validation. Chansol Kim: validation. Jae-Young Choi: methodology. Hyung-Suk Oh and Woong Hee Lee: funding acquisition, conceptualization, supervision, writing. All authors reviewed the manuscript.

Conflicts of interest

There are no conflicts to declare.

Data availability

All supporting data (including experimental spectra, raw images, and LLM results) and the prompts utilized in this paper are provided in the supplementary information (SI) and source data. Source data are provided with this paper. Supplementary information is available. See DOI: https://doi.org/10.1039/d6ta03136f.

Acknowledgements

This work was supported by institutional program grants from the Korea Institute of Science and Technology (KIST) and the

Research Project "Carbon Upcycling Project for Platform Chemicals" of the National Research Foundation (NRF), funded by the Ministry of Science and ICT (grant number: RS-2022-NR068684), Republic of Korea. This research was also supported by a National Research Council of Science & Technology (NST) grant from the Korean Government (MSIT) (No. CAP25011-000 and GTL25021-000).

References

1 A. Urakawa, *Curr. Opin. Chem. Eng.*, 2016, **12**, 31–36.
2 A. Yoon, L. Bai, F. Yang, F. Franco, C. Zhan, M. Ruscher, J. Timoshenko, C. Pratsch, S. Werner, H. S. Jeon, M. C. O. Monteiro, S. W. Chee and B. Roldan Cuenya, *Nat. Mater.*, 2025, **24**, 762–769.
3 C. Zhu, Q. Shi, B. Z. Xu, S. Fu, G. Wan, C. Yang, S. Yao, J. Song, H. Zhou, D. Du, S. P. Beckman, D. Su and Y. Lin, *Adv. Energy Mater.*, 2018, **8**(29), 1801956.
4 J. Wang, C. S. Hsu, T. S. Wu, T. S. Chan, N. T. Suen, J. F. Lee and H. M. Chen, *Nat. Commun.*, 2023, **14**, 6576.
5 Y.-L. Sun, Y.-L. Deng, H.-N. Chen, X.-T. Yang, X.-M. Lin and J.-F. Li, *Small Struct.*, 2023, **4**(6), 2200201.
6 F. Kanoufi, in *Encyclopedia of Electrochemistry*, 2021, pp. 1–80, DOI: 10.1002/9783527610426.bard030108.
7 Y. Wu and N. Liu, *Chem*, 2018, **4**, 438–465.
8 H. Wang, T. Fu, Y. Du, W. Gao, K. Huang, Z. Liu, P. Chandak, S. Liu, P. Van Katwyk, A. Deac, A. Anandkumar, K. Bergen, C. P. Gomes, S. Ho, P. Kohli, J. Lasenby, J. Leskovec, T. Y. Liu, A. Manrai, D. Marks, B. Ramsundar, L. Song, J. Sun, J. Tang, P. Velickovic, M. Welling, L. Zhang, C. W. Coley, Y. Bengio and M. Zitnik, *Nature*, 2023, **620**, 47–60.
9 K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev and A. Walsh, *Nature*, 2018, **559**, 547–555.
10 W. Wang, Y. Liu, Z. Wang, G. Hao and B. Song, *Chem. Sci.*, 2022, **13**, 12604–12615.
11 K. Rajan, H. O. Brinkhaus, M. I. Agea, A. Zielesny and C. Steinbeck, *Nat. Commun.*, 2023, **14**, 5045.
12 A. F. de Almeida, R. Moreira and T. Rodrigues, *Nat. Rev. Chem.*, 2019, **3**, 589–604.
13 E. Chávez-Angel, M. B. Eriksen, A. Castro-Alvarez, J. H. Garcia, M. Botifoll, O. Avalos-Ovando, J. Arbiol and A. Mugarza, *Adv. Intell. Syst.*, 2025, **7**(8), 2400986.
14 E. S. Muckley, R. Vasudevan, B. G. Sumpter, R. C. Advincula and I. N. Ivanov, *ACS Appl. Mater. Interfaces*, 2023, **15**, 2329–2340.
15 Y. Zhang, Y. Han, S. Chen, R. Yu, X. Zhao, X. Liu, K. Zeng, M. Yu, J. Tian, F. Zhu, X. Yang, Y. Jin and Y. Xu, *Nat. Mach. Intell.*, 2025, **7**, 1010–1022.
16 J. Xiong, W. Zhang, Y. Wang, J. Huang, Y. Shi, M. Xu, M. Li, Z. Fu, X. Kong, Y. Wang, Z. Xiong and M. Zheng, *Nat. Mach. Intell.*, 2025, **7**, 782–793.
17 S. Gong, Y. Zhang, Z. Mu, Z. Pu, H. Wang, X. Han, Z. Yu, M. Chen, T. Zheng, Z. Wang, L. Chen, Z. Yang, X. Wu, S. Shi, W. Gao, W. Yan and L. Xiang, *Nat. Mach. Intell.*, 2025, **7**, 543–552.
18 M. B. A, S. Cox, O. Schilter, C. Baldassari, A. D. White and P. Schwaller, *Nat. Mach. Intell.*, 2024, **6**, 525–535.
19 M. Xaba, S. Mapukata and T. Mokhena, *MethodsX*, 2025, **15**, 103525.
20 Z. Zheng, O. Zhang, C. Borgs, J. T. Chayes and O. M. Yaghi, *J. Am. Chem. Soc.*, 2023, **145**, 18048–18062.
21 Z. Zheng, Z. He, O. Khattab, N. Rampal, M. A. Zaharia, C. Borgs, J. T. Chayes and O. M. Yaghi, *Digital Discovery*, 2024, **3**, 491–501.
22 M. Jiang, M. AitTamerd, Z. He, W. Bi, M. Zhao, S. Li and M. Yang, *Mater. Genome Eng. Adv.*, 2026, DOI: 10.1002/mgea.70058.
23 B. Watts, L. Thomsen and P. C. Dastoor, *J. Electron Spectrosc. Relat. Phenom.*, 2006, **151**, 105–120.
24 A. Gaur, B. D. Shrivastava and S. K. Joshi, *J. Phys.: Conf. Ser.*, 2009, **190**, 012084.
25 A. Gaur, B. D. Shrivastava and S. Khalid, *J. Phys.: Conf. Ser.*, 2013, **430**, 012045.
26 C. J. Powell and A. Jablonski, *Nucl. Instrum. Methods Phys. Res. Sect. A Accel. Spectrom. Detect. Assoc. Equip.*, 2009, **601**, 54–65.
27 A. S. Racz and M. Menyhard, *Appl. Surf. Sci. Adv.*, 2025, **30**, 100872.
28 L. A. Ma, F. Massel, A. J. Naylor, L.-C. Duda and R. Younesi, *Commun. Chem.*, 2019, **2**, 125.
29 M. A. van Spronsen, X. Zhao, M. Jaugstetter, C. Escudero, T. Duchon, A. Hunt, I. Waluyo, P. Yang, K. Tschulik and M. B. Salmeron, *J. Phys. Chem. Lett.*, 2021, **12**, 10212–10217.
30 J. Timoshenko and B. Roldan Cuenya, *Chem. Rev.*, 2021, **121**, 882–961.
31 Y. Wang, S. Lany, J. Ghanbaja, Y. Fagot-Revurat, Y. P. Chen, F. Soldera, D. Horwat, F. Mücklich and J. F. Pierson, *Phys. Rev. B*, 2016, **94**, 245418.
32 R. Bunea, A. K. Saikumar and K. Sundaram, *Mater. Sci. Appl.*, 2021, **12**, 315–329.
33 C. Malerba, F. Biccari, C. Leonor Azanza Ricardo, M. D'Incau, P. Scardi and A. Mittiga, *Sol. Energy Mater. Sol. Cells*, 2011, **95**, 2848–2854.
34 H. Ooka, M. C. Figueiredo and M. T. M. Koper, *Langmuir*, 2017, **33**, 9307–9313.
35 M. A. Shumilova, N. N. Pastukhova, N. V. Lomova, I. S. Kazantseva, N. Y. Isupov, D. S. Rybin and F. F. Chausov, *Z. Kristallogr. Cryst. Mater.*, 2025, **240**, 181–189.