



جامعة محمد الخامس بالرباط  
Université Mohammed V de Rabat

**École Nationale Supérieure d'Informatique et d'Analyse des Systèmes**  
Centre d'Études Doctorales en Sciences des Technologies de l'Information et de l'Ingénieur

## **AVIS DE SOUTENANCE DE THÈSE DE DOCTORAT**

**Madame Oumaima REDA**

Soutiendra publiquement sa thèse de Doctorat en Informatique

**Le Mercredi 14 Mai 2025 à 14h00 au Grand Amphi à l'ENSIAS de Rabat**

**Intitulé de la thèse**

# **A New Logic-based Model for Social Media Data Quality Assessment**

**Président :**

Pr. Ali IDRI, PES, ENSIAS, Université Mohammed V de Rabat

**Directeur de thèse :**

Pr. Ahmed ZELLOU, PES, ENSIAS, Université Mohammed V de Rabat

**Rapporteurs :**

Pr Mohammed BERRADA, PES, ENSA, Université Sidi Mohamed Ben Abdellah, Fès

Pr. Hassan SILKAN, PES, Faculté des Sciences, Université Chouaib Doukkali, El Jadida

Pr. Inès CHIHI, MCH, École Nationale d'Ingénieurs de Tunis, Université de Carthage & Faculty of Sciences, Technology and Medicine, University of Luxembourg.

**Examineur :**

Pr. Khalid HOUSNI, MCH, Faculté des Sciences, Université Ibn Tofail, Kénitra



**Résumé:** La croissance rapide d'Internet et les progrès de la technologie de l'information ont considérablement transformé la manière dont les données sont générées, partagées et consommées dans divers domaines. Cette révolution numérique a renforcé le besoin de données pour la prise de décisions éclairées, la création de confiance et un large éventail d'autres applications. Cependant, l'efficacité de ces décisions dépend largement de la qualité de ces données, ce qui rend l'évaluation de la qualité des données un aspect crucial de la gestion des données.

Parmi les sources de données les plus importantes aujourd'hui, les médias sociaux, illustrent parfaitement ce besoin en raison de la complexité, du volume et de la variabilité des données qu'elles produisent. Par conséquent, il est nécessaire d'établir un modèle complet et adaptable pour évaluer la qualité de ces données afin de relever ces défis et garantir une utilisation efficace des données dans diverses applications.

Afin de faire progresser la recherche dans le domaine de la qualité des données, il est important d'avoir une compréhension approfondie de l'état de l'art actuel. Cela implique d'identifier, d'évaluer et d'analyser systématiquement les recherches pertinentes menées ces dernières années. Pour ce faire, une revue systématique de la littérature (SLR) a été réalisée comme approche appropriée pour explorer le contexte de la qualité des données et examiner les recherches existantes. S'appuyant sur les connaissances obtenues lors de cette revue, cette thèse aborde les défis identifiés et contribue au domaine de la qualité des données via trois contributions principales : La première vise à proposer une nouvelle classification des dimensions de la qualité des données basée sur les modèles, en vue de définir un nouveau modèle de qualité des données des médias sociaux appelé SMDQM, afin d'identifier les dimensions les plus couramment utilisées pour évaluer de manière fiable la qualité des données des médias sociaux. La deuxième contribution introduit un nouveau modèle de qualité efficace, FULMQA, pour l'évaluation de la qualité des données des médias sociaux, comprenant huit métriques de qualité nécessaires pour une évaluation précise. Enfin, la troisième contribution étend le modèle FULMQA en proposant un modèle robuste à deux niveaux (2L-DQM) qui évalue la qualité des données des médias sociaux à la fois au niveau du contenu et des utilisateurs. Ce modèle intègre diverses métriques, y compris des indicateurs quantitatifs et qualitatifs, afin de fournir une analyse plus complète, capturant la nature multidimensionnelle des interactions sur les médias sociaux.

Afin de traiter l'imprécision et la subjectivité inhérentes aux évaluations de la qualité des données, nous avons modélisé nos métriques à l'aide de la logique floue, permettant l'utilisation d'expressions linguistiques dans le processus d'évaluation. Des évaluations expérimentales approfondies ont été réalisées à l'aide de deux ensembles de données réelles. Les résultats montrent que nos modèles sont très efficaces pour évaluer les données des médias sociaux, obtenant de solides performances sur toutes les données testées.

**Mots-clés:** Qualité des données, Métriques de qualité, Modèle de qualité, Évaluation de la qualité, Médias sociaux.



**Abstract:** The fast growth of the Internet and advances in information technology have significantly changed the way data is generated, shared, and consumed across various domains. This digital revolution has reinforced the need for data-driven insights for informed decision-making, fostering trust, and a wide range of other applications. However, the effectiveness of such decisions is highly dependent on the quality of these data, making data quality assessment a crucial aspect of data management.

Among the most prominent data sources today, social media platforms exemplify this need given the complexity, volume, and variability of the data they produce. Consequently, establishing a comprehensive and adaptable model for evaluating the quality of this data is necessary to address these challenges and ensure the effective use of data in diverse applications.

To advance research in the field of data quality, it is important to have a comprehensive understanding of the current state of the art. This involves systematically identifying, evaluating, and analyzing relevant research conducted in recent years. To achieve this, a systematic literature review (SLR) was conducted as an appropriate approach to explore the landscape of data quality and examine existing research focused on both general data quality assessment (DQA) and the specific context of social media data quality. Building on the insights derived from the SLR, this thesis addresses identified challenges and contributes to the field in three main ways:

The first main contribution aims to provide a new data quality dimension classification based on the existing quality models collected during the SLR, towards a social media data quality model called SMDQM to identify the most commonly used dimensions needed to assess the quality of social media data reliably. As a second contribution, we introduce a novel and efficient quality model, FULMQA, for the assessment of social media data quality, including eight quality metrics needed for an accurate assessment. Our third contribution extends the FULMQA, proposing a robust two-level model (2L-DQM) that assesses the quality of social media data at both the content and user levels. This model integrates diverse metrics, including both quantitative and qualitative indicators, to provide a more comprehensive analysis, capturing the multidimensional nature of social media interactions.

To address the inherent imprecision and subjectivity in data quality assessments, we model our metrics using fuzzy logic, allowing the use of linguistic expressions in the evaluation process. Extensive experimental evaluations were performed using two real-world datasets. The results demonstrate that our models are highly effective in evaluating social media data, achieving strong performance across all tested datasets.

**Keywords:** Data Quality, Quality Metrics, Quality Model, Quality Assessment, Social Media