Le web scraping, parfois appelé harvesting ou en français moissonnage, est une technique de « récupération et organisation automatisées des données Web » ; c'est la principale forme de data mining et d'extraction des données de sites web, via un script ou un programme. Il vise à capter des données, pour les transformer et/ou les réutiliser dans un autre contexte comme l'enrichissement de bases de données, le référencement ou l'exploration de données, ou l'apprentissage profond pour une intelligence artificielle. L'objectif est souvent commercial, mais parfois scientifique ou politique.  
Profitant des progrès logiciels et de la science des données, dans le cadre de l'économie numérique où au sein de l'économie de la surveillance, l'exploitation commercial et politique du big data le moissonnage du Web s'est exponentiellement généralisé, illégalement parfois, ou en bénéficiant de flous ou vides juridiques.  
Les enjeux sont notamment commerciaux et politiques, mais aussi éthiques, juridiques et philosophiques, car le scraping met en péril la protection des données et la sécurité des individus et de certains groupes. Ces enjeux deviennent critiques avec l'émergence des intelligences artificielles (génératives et multimodales notamment) qui permettent d'utiliser à mauvais escient des données volées et/ou apparemment banales et sans intérêt massivement scrapées sur l'Internet pour des activités criminelles ou pour la propagande, y compris électorale, comme l'ont montré le scandale Facebook-Cambridge Analytica/AggregateIQ et la diffusion croissante de fake news ciblées. Parmi les enjeux émergents figurent l'utilisation de l'intelligence artificielle qui à la fois facilite le moissonnage et permet d'en tirer beaucoup plus d'informations, y compris des informations sensibles et sur la vie privée. « Aujourd’hui, les chercheurs s'appuient de plus en plus sur l'intelligence artificielle (par exemple, l'apprentissage automatique, l'analyse de sentiments et/ou l'apprentissage profond) pour extraire le sens et l’interférence des données Web »).  
Le Web scraping comme source de préoccupation  
Les outils récents de « moissonnage des données » (originelles ou dupliquées) permettent d'extraire et collecter, puis traiter (classer, vérifier, nettoyer, agréger et structurer) la donnée, généralement à partir « Big Web Data » dont le volume croît exponentiellement (multiplié par plus de trente en 10 ans, passant de 2 zettaoctets en 2010 à 64 zettaoctets en 2023 . Cette donnée est le plus souvent directement captée dans le code source de pages web (actuelles ou archivées) et dans les plateformes du Web (réseaux sociaux notamment). Selon Statista, en 2020, environ 64 000 milliards de gigaoctets de données ont été générées, dont seul un petit pourcentage sera conservé car la capacité de stockage mondiale n'était "que" de 6,7 zettaoctets en 2020 (environ 6,7 milliards de gigaoctets). le volume de la donnée du Web devrait cependant dépasser 180 zettaoctets vers 2025, à cause du développement des objets connectés et de la 5G.

Avec une efficacité et une rapidité croissante (liée aux progrès de l'informatique), ces outils permettent, à grande échelle, d'extraire, structurer, analyser et évaluer, puis stocker et manipuler, et éventuellement commercialiser, ces données et des données dérivées (secondaires). Les bases de données ainsi produites peuvent inclure un grand nombre de renseignements personnels et des données sur toutes sortes d'entités socioéconomiques, culturelles, médiatiques, militaires, etc. Elles permettent de réaliser des profils psychologiques, sociaux et économiques très poussés de centaines de milliards de personnes, ainsi que des profils d'entreprises, d'ONG, etc.  
Ces données, anonymisées, ont d'abord été utilisées pour diverses études scientifiques (en épidémiologie et en sociologie et anthropologie par exemple), mais dans un monde de plus en plus axé sur l'économie de l'information, ces données sont maintenant devenues une « ressource stratégique comparable en valeur à la terre, à l'or et au pétrole »,, très recherchées par des officines ou entreprises spécialisées qui vendent ces données agrégées aux banques, aux assureurs, à certains groupes politiques, des États autoritaires et dictatures (qui les utilisent pour cibler et contrôler des citoyens et diverses entités).  
Alors que les algorithmes sophistiqués et l'intelligence artificielle peuvent aisément désanonymiser un grand nombre de données personnelles (données de santé y compris), le moissonnage massif et permanent des données du Web, par les gouvernements et par un nombre croissant de plateformes Web privées et d'officines dont le commerce de la donnée fonde le modèle économique soulève de graves préoccupations éthiques et politiques en matière de protection de la vie privée et de données sensibles.  
Parmi les dérives souvent citées figurent l'espionnage massif des citoyens (y compris les leurs), et des concurrents économiques par les États-Unis, la Russie et la Chine, notamment.  
Les systèmes américains d'espionnage généralisé (mis au jour par divers lanceurs d'alerte dont Julian Assange, Chelsea Manning, Edward Snowden), les fermes de Trolls russes ou l'approche chinoise du système de crédit social (SCS), « le premier système de notation sociale mis en œuvre numériquement à l'échelle nationale, vise à améliorer la fiabilité au sein de la société chinoise et constitue un exemple essentiel de la transformation numérique de la société Il est en outre conçu pour améliorer le comportement moral, la fiabilité financière et le contrôle social ».  
Le scraping du web menace les quatre droits et « états de base de la vie privée » énoncés par Westin en 1962 (droit à la solitude, à l'intimité, à l'anonymat dans la foule et à la réserve) qui sont menacés dans un nombre croissant de situations, de même que la protection du contenu des courriers électroniques qui fait partie du droit à la vie privée.  
Légalité  
Le scrapping doit respecter la propriété commerciale et/ou intellectuelle et, dans l'Union européenne, le RGPD. Sans autorisation préalable, il n'est autorisé que pour des données nativement légalement publiques, ou tombées dans le domaine public, ou pour des données stockées sur des sites ou plateformes sous une licence libre le permettant (certaines licences ouvertes (Creative Commons Attribution par exemple) stipulent que le droit de réutilisation et/ou de transformation de la donnée ne vaut que pour les usages non commerciaux et/ou que les nouvelles donnés soient également publiées sous une licence ouverte similaire, en mentionnant l'auteur initial, la source, la licence, etc.).